Large Deviations

Kun Dong, Daniel Freund, Matt Hin,
Andrew Loeb, and Aditya Vaidyanathan

Cornell University

November 18, 2014

Overview

Setting

Let $X_1, X_2, \ldots$ be i.i.d., $S_n = X_1 + \ldots + X_n$.

We investigate the rate at which $P(S_n \geq na) \to 0$ for $a > \mu = E[X_i] < \infty$.

Setting

Let $X_1, X_2, \ldots$ be i.i.d., $S_n = X_1 + \ldots + X_n$.

We investigate the rate at which $P(S_n \geq na) \to 0$ for $a > \mu = E[X_i] < \infty$.

We define the *moment-generating function* $\phi(\theta) = E[e^{\theta X_i}]$.

Setting

Let $X_1, X_2, \ldots$ be i.i.d., $S_n = X_1 + \ldots + X_n$.

We investigate the rate at which $P(S_n \geq na) \to 0$ for $a > \mu = E[X_i] < \infty$.

We define the *moment-generating function* $\phi(\theta) = E[e^{\theta X_i}]$.

Eventually, we will see that if $\phi(\theta) < \infty$ for some $\theta > 0$, then $P(S_n \geq na) \to 0$ exponentially fast.

Setting

Let $X_1, X_2, \ldots$ be i.i.d., $S_n = X_1 + \ldots + X_n$.

We investigate the rate at which $P(S_n \geq na) \to 0$ for $a > \mu = E[X_i] < \infty$.

We define the *moment-generating function* $\phi(\theta) = E[e^{\theta X_i}]$.

Eventually, we will see that if $\phi(\theta) < \infty$ for some $\theta > 0$, then $P(S_n \geq na) \to 0$ exponentially fast.

SLLN? CLT? Chebyshev?

Convergence rate $\gamma(a)$ - Part I

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

Convergence rate $\gamma(a)$ - Part I

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

We first need to show that γ exists. Let $\pi_n = P(S_n \geq na)$.

Convergence rate $\gamma(a)$ - Part I

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

We first need to show that γ exists. Let $\pi_n = P(S_n \geq na)$. Then,

$$\pi_{m+n} = P(S_{n+m} \geq (n+m)a) \geq P(S_m \geq ma,\ S_{n+m} - S_m \geq na) = \pi_m \pi_n,$$

Convergence rate $\gamma(a)$ - Part I

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

We first need to show that γ exists. Let $\pi_n = P(S_n \geq na)$. Then,

$$\pi_{m+n} = P(S_{n+m} \geq (n+m)a) \geq P(S_m \geq ma,\ S_{n+m} - S_m \geq na) = \pi_m \pi_n,$$

where the inequality holds, because

$$\begin{aligned} & S_m \geq ma,\ S_{n+m} - S_m \geq na \\ & \quad \implies S_{n+m} = S_m + S_{n+m} - S_m \geq ma + na = a(m+n) \end{aligned}$$

Convergence rate $\gamma(a)$ - Part I

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na)$$

We first need to show that γ exists. Let $\pi_n = P(S_n \geq na)$. Then,

$$\pi_{m+n} = P(S_{n+m} \geq (n+m)a) \geq P(S_m \geq ma,\ S_{n+m} - S_m \geq na) = \pi_m \pi_n,$$

where the inequality holds, because

$$S_m \geq ma,\ S_{n+m} - S_m \geq na \implies S_{n+m} = S_m + S_{n+m} - S_m \geq ma + na = a(m+n)$$

and the equality because $S_m = X_1 + \ldots X_m$ and $S_{n+m} - S_m = X_{m+1} + \ldots X_{n+m}$ are independent.

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Lemma: $\tau_{m+n} \geq \tau_m + \tau_n \implies \frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$ as $n \to \infty$.

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Lemma: $\tau_{m+n} \geq \tau_m + \tau_n \implies \frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$ as $n \to \infty$.

Proof: $\limsup \frac{\tau_n}{n} \leq \sup \frac{\tau_m}{m}$ gives $\leq$.

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Lemma: $\tau_{m+n} \geq \tau_m + \tau_n \implies \frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$ as $n \to \infty$.

Proof: $\limsup \frac{\tau_n}{n} \leq \sup \frac{\tau_m}{m}$ gives $\leq$.

If $\forall m : \liminf \frac{\tau_n}{n} \geq \frac{\tau_m}{m}$, then $\geq$ holds as well.

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Lemma: $\tau_{m+n} \geq \tau_m + \tau_n \implies \frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$ as $n \to \infty$.

Proof: $\limsup \frac{\tau_n}{n} \leq \sup \frac{\tau_m}{m}$ gives $\leq$.

If $\forall m : \liminf \frac{\tau_n}{n} \geq \frac{\tau_m}{m}$, then $\geq$ holds as well.

Write $n = km + l$, where $0 \leq l < m$. Then

$$\tau_n = \tau_{km+l} \geq \tau_m + \tau_{(k-1)m+l} \geq \dots \geq k\tau_m + \tau_l.$$

Convergence rate $\gamma(a)$ - Part II

Set $\tau_n = \log(\pi_n)$.

Lemma: $\tau_{m+n} \geq \tau_m + \tau_n \implies \frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$ as $n \to \infty$.

Proof: $\limsup \frac{\tau_n}{n} \leq \sup \frac{\tau_m}{m}$ gives $\leq$.

If $\forall m : \liminf \frac{\tau_n}{n} \geq \frac{\tau_m}{m}$, then $\geq$ holds as well.

Write $n = km + l$, where $0 \leq l < m$. Then

$$\tau_n = \tau_{km+l} \geq \tau_m + \tau_{(k-1)m+l} \geq \ldots \geq k\tau_m + \tau_l.$$

If we now divide both sides by $n = km + l$, we find that

$$\frac{\tau_n}{n} \geq \left(\frac{k}{km+l}\right)\tau_m + \frac{\tau_l}{n} = \left(\frac{km}{km+l}\right)\frac{\tau_m}{m} + \frac{\tau_l}{n} \to \frac{\tau_m}{m}. \quad \square$$

Convergence rate $\gamma(a)$ - Part III

We have now seen that $\pi_{n+m} \geq \pi_m \pi_n$, implying that

Convergence rate $\gamma(a)$ - Part III

We have now seen that $\pi_{n+m} \geq \pi_m \pi_n$, implying that

$$\tau_{n+m} = \log(\pi_{n+m}) \geq \log(\pi_m \pi_n) = \log(\pi_m) + \log(\pi_n) = \tau_m + \tau_n$$

Convergence rate $\gamma(a)$ - Part III

We have now seen that $\pi_{n+m} \geq \pi_m \pi_n$, implying that

$$\tau_{n+m} = \log(\pi_{n+m}) \geq \log(\pi_m \pi_n) = \log(\pi_m) + \log(\pi_n) = \tau_m + \tau_n$$

But we have also seen that $\tau_{n+m} \geq \tau_m + \tau_n$ implies that $\frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$. Since $\frac{\tau_n}{n} = \frac{1}{n} \log P(S_n \geq na)$,

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na) = \sup_m \frac{\log(P(S_m \geq ma))}{m} \leq 0 \text{ exists.}$$

Convergence rate $\gamma(a)$ - Part III

We have now seen that $\pi_{n+m} \geq \pi_m \pi_n$, implying that

$$\tau_{n+m} = \log(\pi_{n+m}) \geq \log(\pi_m \pi_n) = \log(\pi_m) + \log(\pi_n) = \tau_m + \tau_n$$

But we have also seen that $\tau_{n+m} \geq \tau_m + \tau_n$ implies that $\frac{\tau_n}{n} \to \sup_m \frac{\tau_m}{m}$. Since $\frac{\tau_n}{n} = \frac{1}{n} \log P(S_n \geq na)$,

$$\gamma(a) = \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na) = \sup_m \frac{\log(P(S_m \geq ma))}{m} \leq 0 \text{ exists.}$$

Then, since $\frac{\log P(S_n \geq na)}{n} \leq \sup_m \frac{\log(P(S_m \geq ma))}{m} = \gamma(a)$,

$$P(S_n \geq na) \leq e^{\gamma(a)n}.$$

Understanding $\gamma(a)$ better

$$\gamma(a) = -\infty \Leftrightarrow P(X_1 \geq a) = 0 \Leftrightarrow P(S_n \geq na) = 0 \ \forall n$$

Understanding $\gamma(a)$ better

$$\gamma(a) = -\infty \Leftrightarrow P(X_1 \geq a) = 0 \Leftrightarrow P(S_n \geq na) = 0 \; \forall n$$

$\gamma(a) = -\infty \implies P(S_n \geq na) = 0 \; \forall n$: note that

$\gamma(a) \geq \frac{\log(P(S_n \geq na))}{n} \forall n$, so $\log(P(S_n \geq na)) = -\infty$

$P(S_n \geq na) = 0 \; \forall n \implies P(X_i \geq a) = 0$:
set $n = 1$, X_i are i.i.d.

$P(X_i \geq a) = 0 \implies \gamma(a) = -\infty$:
$P(X_i \geq a) = 0$, so $\forall n : P(S_n \geq na) = 0$, so $\lim_{n \to \infty} \frac{\log(0)}{n} = -\infty$.

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na} P(S_n \geq na) \leq E e^{\theta S_n}$$

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na} P(S_n \geq na) \leq E e^{\theta S_n} = E\left(\prod_{i=1}^{n} e^{\theta X_i}\right)$$

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na}P(S_n \geq na) \leq Ee^{\theta S_n} = E\left(\prod_{i=1}^{n} e^{\theta X_i}\right) = \prod_{i=1}^{n} Ee^{\theta X_i}$$

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na} P(S_n \geq na) \leq E e^{\theta S_n} = E\left(\prod_{i=1}^{n} e^{\theta X_i}\right) = \prod_{i=1}^{n} E e^{\theta X_i} = (\phi(\theta))^n$$

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na} P(S_n \geq na) \leq E e^{\theta S_n} = E\left(\prod_{i=1}^{n} e^{\theta X_i}\right) = \prod_{i=1}^{n} E e^{\theta X_i} = (\phi(\theta))^n$$

Rearranging,

$$e^{\theta na} P(S_n \geq na) \leq (\phi(\theta))^n \implies P(S_n \geq na) \leq \exp\left[-n(a\theta - \kappa(\theta))\right]$$

for $\kappa(\theta) = \log \phi(\theta)$.

Bound for $\gamma(a)$

We seek to bound $\gamma(a)$. By Chebyshev's inequality:

$$e^{\theta na} P(S_n \geq na) \leq Ee^{\theta S_n} = E\left(\prod_{i=1}^{n} e^{\theta X_i}\right) = \prod_{i=1}^{n} Ee^{\theta X_i} = (\phi(\theta))^n$$

Rearranging,

$$e^{\theta na} P(S_n \geq na) \leq (\phi(\theta))^n \implies P(S_n \geq na) \leq \exp\left[-n(a\theta - \kappa(\theta))\right]$$

for $\kappa(\theta) = \log \phi(\theta)$. Hence,

$$\gamma(a) \leq -\{a\theta - \kappa(\theta)\}$$

for some fixed θ. We want $a\theta - \kappa(\theta) > 0$.

Moment Generating Functions

Definition

The *moment generating function* (MGF) for a random variable X is defined to be $\phi(\theta) = Ee^{\theta X}$.

Moment Generating Functions

Definition

The *moment generating function* (MGF) for a random variable X is defined to be $\phi(\theta) = Ee^{\theta X}$.

MGFs vs. Characteristic Functions

- Characteristic functions can be thought of as the MGF of iX
- Characteristic functions always exist and is complex valued.
- MGFs does not necessarily exist for all θ and is real valued.

Moment Generating Functions

Definition

The *moment generating function* (MGF) for a random variable X is defined to be $\phi(\theta) = Ee^{\theta X}$.

MGFs vs. Characteristic Functions

- Characteristic functions can be thought of as the MGF of iX
- Characteristic functions always exist and is complex valued.
- MGFs does not necessarily exist for all θ and is real valued.

Generating Moments with MGFs

$$\left(\frac{d}{d\theta}\int_{-\infty}^{\infty} e^{\theta x}\, dF(x)\right)\Bigg|_{\theta=0} = \left(\int_{-\infty}^{\infty} x^n e^{\theta x}\, dF(x)\right)\Bigg|_{\theta=0} = EX^N.$$

Generating Moments with MGFs

Example: $X \sim \text{Unif}(0,1)$

Generating Moments with MGFs

Example: $X \sim \text{Unif}(0,1)$

$$\phi(\theta) = Ee^{\theta X} = \int_0^1 e^{\theta x}\, dx = \frac{e^{\theta} - 1}{\theta}.$$

Generating Moments with MGFs

Example: $X \sim \text{Unif}(0,1)$

$$\phi(\theta) = Ee^{\theta X} = \int_0^1 e^{\theta x}\, dx = \frac{e^\theta - 1}{\theta}.$$

We obtain moments by differentiating and evaluating at $\theta = 0$:

$$EX = \frac{d}{d\theta}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{e^\theta(\theta - 1) + 1}{\theta^2}\Big|_0 = \frac{1}{2},$$

Generating Moments with MGFs

Example: $X \sim \text{Unif}(0,1)$

$$\phi(\theta) = Ee^{\theta X} = \int_0^1 e^{\theta x}\,dx = \frac{e^\theta - 1}{\theta}.$$

We obtain moments by differentiating and evaluating at $\theta = 0$:

$$EX = \frac{d}{d\theta}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{e^\theta(\theta - 1) + 1}{\theta^2}\Big|_0 = \frac{1}{2},$$

$$EX^2 = \frac{d^2}{d\theta^2}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{e^\theta(\theta^2 - 2\theta + 2) - 2}{\theta^3}\Big|_0 = \frac{1}{3},$$

Generating Moments with MGFs

Example: $X \sim \text{Unif}(0,1)$

$$\phi(\theta) = Ee^{\theta X} = \int_0^1 e^{\theta x}\,dx = \frac{e^\theta - 1}{\theta}.$$

We obtain moments by differentiating and evaluating at $\theta = 0$:

$$EX = \frac{d}{d\theta}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{e^\theta(\theta - 1) + 1}{\theta^2}\Big|_0 = \frac{1}{2},$$

$$EX^2 = \frac{d^2}{d\theta^2}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{e^\theta(\theta^2 - 2\theta + 2) - 2}{\theta^3}\Big|_0 = \frac{1}{3},$$

$$EX^n = \frac{d^n}{d\theta^n}\left(\frac{e^\theta - 1}{\theta}\right)\Big|_0 = \frac{1}{n+1}.$$

Restriction on the Distribution of X_i

Assumption H1

The moment generating function $\phi(\theta) = Ee^{\theta X_i} < \infty$ for some $\theta > 0$.

Restriction on the Distribution of X_i

Assumption H1

The moment generating function $\phi(\theta) = Ee^{\theta X_i} < \infty$ for some $\theta > 0$.

Note that $\phi(0) = 1$ for any distribution and let

$$\theta_- = \inf\{\theta : \phi(\theta) < \infty\}$$
$$\theta_+ = \sup\{\theta : \phi(\theta) < \infty\}$$

Restriction on the Distribution of X_i

Assumption H1

The moment generating function $\phi(\theta) = Ee^{\theta X_i} < \infty$ for some $\theta > 0$.

Note that $\phi(0) = 1$ for any distribution and let

$$\theta_- = \inf\{\theta : \phi(\theta) < \infty\} \leq 0,$$
$$\theta_+ = \sup\{\theta : \phi(\theta) < \infty\} > 0.$$

Restriction on the Distribution of X_i

Assumption H1

The moment generating function $\phi(\theta) = Ee^{\theta X_i} < \infty$ for some $\theta > 0$.

Note that $\phi(0) = 1$ for any distribution and let

$$\theta_- = \inf\{\theta : \phi(\theta) < \infty\} \leq 0,$$
$$\theta_+ = \sup\{\theta : \phi(\theta) < \infty\} > 0.$$

Remarks

If $\theta_- < 0$, then

- moments of all orders are finite, and
- the tails of the distributions of X_i are exponentially bounded.

(H1) Implies $\mu = EX_i \neq \infty$

Proof.

Let $F(x) = P(X_i \leq x)$ and fix θ from (H1). If $\theta \geq \frac{1}{e}$, (H1) directly bounds μ. Else, then $x \geq e^{\theta x}$ in the interval (r_1, r_2). Hence:

$$\begin{aligned} EX_i \leq EX_i^+ &= \int_0^\infty x\, dF(x), \\ &\leq \int_{r_1}^{r_2} x\, dF(x) + \int_0^\infty e^{\theta x}\, dF(x), \\ &\leq r_2 + \phi(\theta), \\ &< \infty. \end{aligned}$$

□

Good check that $a > \mu$ is sound.

Decay Bound

Lemma (Exponential Decay Bound)

If $a > \mu$ and $\theta > 0$ is small, then $a\theta - \kappa(\theta) > 0$.

Motivation

$a\theta - \kappa(\theta) = \int_0^\theta (a - \kappa'(x))\, dx$ and $\kappa(0) = \log \phi(0) = 0$.

(1) $\kappa(\theta)$ is continuous at $\theta = 0$.

(2) κ is differentiable over $(0, \theta_+)$.

(3) $\kappa'(\theta) \to \mu$ as $\theta \to 0$.

So there exists some $\theta_0 > 0$ such that $a\theta - \kappa(\theta) > 0$ for $\theta \in (0, \theta_0)$.

Decay Bound Proof - Part I

Condition (1)

Let $F(x) = P(X_i \leq x)$. For $0 < \theta < \theta_0 < \theta_+$, then we can dominate $e^{\theta x} \leq 1 + e^{\theta_0 x}$. By the DCT:

$$\lim_{\theta \to 0} \int e^{\theta x} \, dF(x) = \int dF(x) = 1.$$

This implies that $\phi(\theta)$ is continuous at $\theta = 0$ hence $\kappa(\theta)$ is continuous at $\theta = 0$.

Decay Bound Proof - Part II

Condition (2)

For $|h| < h_0$, then $\left|e^{hx} - 1\right| = \left|\int_0^{hx} e^y \, dy\right| \leq |hx| e^{h_0 x}$. Consider,

$$
\begin{aligned}
\phi'(\theta) &= \lim_{h \to 0} \frac{\phi(\theta + h) - \phi(\theta)}{h}, \\
&= \lim_{h \to 0} \int \frac{e^{hx} - 1}{h} \cdot e^{\theta x} \, dF(x), \\
&= \int x e^{\theta x} \, dF(x), \quad \text{for } \theta \in (0, \theta_+).
\end{aligned}
$$

Hence, $\kappa'(\theta) = \frac{\phi'(\theta)}{\phi(\theta)}$ exists for $\theta \in (0, \theta_+)$.

Decay Bound Proof - Part III

Condition (3)

Note that we will use the DCT with the inequality:

$$e^{\theta x} \leq 1 + e^{\theta_0 x}.$$

Hence,

$$\phi(0) = \lim_{\theta \to 0} \int e^{\theta x} \, dF(x) = \int \left(\lim_{\theta \to 0} e^{\theta x} \right) dF(x) = \int dF(x) = 1,$$

$$\phi'(0) = \lim_{\theta \to 0} \int x e^{\theta x} \, dF(x) = \int x \left(\lim_{\theta \to 0} e^{\theta x} \right) dF(x) = \int x \, dF(x) = \mu.$$

So, $\kappa'(\theta) \to \mu$ as $\theta \to 0$. □

An Upper Bound

We just showed there exists $\theta_0 \in (0, \theta_+)$ such that

$$a\,\theta - \kappa(\theta) > 0 \qquad \text{for} \qquad \theta \in (0, \theta_0)$$

Earlier we have shown that for any $\theta \in (0, \theta_+)$,

$$P(S_n \geq na) \leq exp(-n\{a\,\theta - \kappa(\theta)\})$$

where $\kappa(\theta) = \log \phi(\theta)$.

An Upper Bound

We just showed there exists $\theta_0 \in (0, \theta_+)$ such that

$$a\,\theta - \kappa(\theta) > 0 \qquad \text{for} \qquad \theta \in (0, \theta_0)$$

Earlier we have shown that for any $\theta \in (0, \theta_+)$,

$$P(S_n \geq na) \leq exp(-n\{a\,\theta - \kappa(\theta)\})$$

where $\kappa(\theta) = \log \phi(\theta)$.

An Upper Bound

We just showed there exists $\theta_0 \in (0, \theta_+)$ such that

$$a\,\theta - \kappa(\theta) > 0 \qquad \text{for} \qquad \theta \in (0, \theta_0)$$

Earlier we have shown that for any $\theta \in (0, \theta_+)$,

$$P(S_n \geq na) \leq exp(-n\{a\,\theta - \kappa(\theta)\})$$

where $\kappa(\theta) = \log \phi(\theta)$.

$$\Longrightarrow \lim_{n \to \infty} \frac{1}{n} \log P(S_n \geq na) \leq -\{a\,\theta - \kappa(\theta)\}$$

An Upper Bound

We just showed there exists $\theta_0 \in (0, \theta_+)$ such that

$$a\,\theta - \kappa(\theta) > 0 \qquad \text{for} \qquad \theta \in (0, \theta_0)$$

Earlier we have shown that for any $\theta \in (0, \theta_+)$,

$$P(S_n \geq na) \leq exp(-n\{a\,\theta - \kappa(\theta)\})$$

where $\kappa(\theta) = \log \phi(\theta)$.

$$\Longrightarrow \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na) \leq -\{a\,\theta - \kappa(\theta)\}$$

The expression in the braces gives an upper bound on the rate of the exponential decay whenever it is positive.

An Upper Bound

We just showed there exists $\theta_0 \in (0, \theta_+)$ such that

$$a\,\theta - \kappa(\theta) > 0 \qquad \text{for} \qquad \theta \in (0, \theta_0)$$

Earlier we have shown that for any $\theta \in (0, \theta_+)$,

$$P(S_n \geq na) \leq exp(-n\{a\,\theta - \kappa(\theta)\})$$

where $\kappa(\theta) = \log \phi(\theta)$.

$$\Longrightarrow \lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na) \leq -\{a\,\theta - \kappa(\theta)\}$$

The expression in the braces gives an upper bound on the rate of the exponential decay whenever it is positive.
Each feasible θ gives such a bound, so it is natural to find out the best bound by maximizing $a\,\theta - \kappa(\theta)$ over $(0, \theta_+)$.

When Things Are Nice

We find the maximum of $\theta a - \kappa(\theta)$ at its critical point

$$\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$$

The maximum occurs when $a = \phi'(\theta)/\phi(\theta)$.

When Things Are Nice

We find the maximum of $\theta a - \kappa(\theta)$ at its critical point

$$\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$$

The maximum occurs when $a = \phi'(\theta)/\phi(\theta)$.
Assumptions we need to make:

- There exists exactly one critical point.
- The critical point is a maximum.

Justify the Assumptions

For any $\phi(\theta) < \infty$, define

$$F_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{x} e^{\theta y} \, dF(y)$$

Justify the Assumptions

For any $\phi(\theta) < \infty$, define

$$F_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{x} e^{\theta y}\, dF(y)$$

Claim: F_θ is a distribution function for $\theta \in (\theta_-, \theta_+)$.

Justify the Assumptions

For any $\phi(\theta) < \infty$, define

$$F_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{x} e^{\theta y}\, dF(y)$$

Claim: F_θ is a distribution function for $\theta \in (\theta_-, \theta_+)$.
Proof:

$$F_\theta(x) = \frac{1}{\phi(\theta)} \int 1_{(-\infty, x]}(y)\, e^{\theta y}\, dF(y)$$

By Dominated Convergence Theorem, $F_\theta(-\infty) = 0$ and $F_\theta(\infty) = 1$. It is non-decreasing because $e^{\theta y}$ is non-negative. It is right-continuous because

$$|F_\theta(x+\epsilon) - F_\theta(x)| = \frac{1}{\phi(\theta)} \int 1_{(x, x+\epsilon]}(y)\, e^{\theta y}\, dF(y) \downarrow 0$$

again by Dominated Convergence Theorem. □

Distribution F_θ

We can compute the mean of distribution function F_θ we just defined.

Distribution F_θ

We can compute the mean of distribution function F_θ we just defined.

$$\int x\, dF_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{\infty} x\, e^{\theta x}\, dF(x) = \frac{\phi'(\theta)}{\phi(\theta)} = \kappa'(\theta)$$

Distribution F_θ

We can compute the mean of distribution function F_θ we just defined.

$$\int x\, dF_\theta(x) = \frac{1}{\phi(\theta)} \int_{-\infty}^{\infty} x\, e^{\theta x}\, dF(x) = \frac{\phi'(\theta)}{\phi(\theta)} = \kappa'(\theta)$$

Outline of proof:

The second equality was proven in the previous section.
Let μ be the (Lebesgue-Stieltjes) measure induced by F_θ, we have $\mu((a,b]) = \frac{1}{\phi(\theta)} \int 1_{(a,b]}(y)\, e^{\theta y}\, dF(y)$ from definition.
(1) The collection of sets with the property

$$\mu(E) = \frac{1}{\phi(\theta)} \int 1_E(y)\, e^{\theta y}\, dF(y)$$

forms a σ-algebra, so it includes the Borel sets.
(2) For general measurable function g,

$$\int g(x)\, dF_\theta(x) = \frac{1}{\phi(\theta)} \int g(y)\, e^{\theta y}\, dF(y).$$

In particular, we can let $g(x) = x$. □

$\phi''(\theta)$

$$
\begin{aligned}
\phi''(\theta) &= \lim_{h\to 0} \frac{\phi'(\theta+h)-\phi'(\theta)}{h} \\
&= \lim_{h\to 0} \int \frac{e^{hx}-1}{h} x\, e^{\theta x}\, dF(x) \\
&\overset{DCT}{=} \int x^2\, e^{\theta x}\, dF(x) = \phi(\theta) \int x^2\, dF_\theta(x)
\end{aligned}
$$

$\phi''(\theta)$

$$\phi''(\theta) = \lim_{h\to 0} \frac{\phi'(\theta+h) - \phi'(\theta)}{h}$$

$$= \lim_{h\to 0} \int \frac{e^{hx} - 1}{h} x\, e^{\theta x}\, dF(x)$$

$$\overset{DCT}{=} \int x^2\, e^{\theta x}\, dF(x) = \phi(\theta) \int x^2\, dF_\theta(x)$$

To apply the dominated convergence theorem, we fix small h_0 and ϵ. For $h < h_0$

$$\left| \frac{e^{hx} - 1}{h} x\, e^{\theta x} \right| \le |x^2| e^{(\theta+h_0)x} \le \frac{2}{\epsilon^2} e^{\epsilon|x|} e^{(\theta+h_0)x}$$

$$\le \frac{2}{\epsilon^2} (e^{(\theta+h_0+\epsilon)x} + e^{(\theta+h_0-\epsilon)x})$$

Second Derivative Test

Recall that the function we are trying to maximize is $a\theta - \kappa(\theta)$.

Second Derivative Test

Recall that the function we are trying to maximize is $a\,\theta - \kappa(\theta)$. Its derivative is $\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$.

Second Derivative Test

Recall that the function we are trying to maximize is $a\,\theta - \kappa(\theta)$.
Its derivative is $\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$.
Its second derivative is $-\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)}$.

Second Derivative Test

Recall that the function we are trying to maximize is $a\,\theta - \kappa(\theta)$.
Its derivative is $\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$.
Its second derivative is $-\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)}$.

$$\begin{aligned}\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} &= \frac{\phi''(\theta)}{\phi(\theta)} - \left(\frac{\phi'(\theta)}{\phi(\theta)}\right)^2 \\ &= \int x^2\,dF_\theta(x) - \left(\int x\,dF_\theta(x)\right)^2 \geq 0\end{aligned}$$

Second Derivative Test

Recall that the function we are trying to maximize is $a\,\theta - \kappa(\theta)$.
Its derivative is $\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$.
Its second derivative is $-\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)}$.

$$\begin{aligned}\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} &= \frac{\phi''(\theta)}{\phi(\theta)} - \left(\frac{\phi'(\theta)}{\phi(\theta)}\right)^2 \\ &= \int x^2\,dF_\theta(x) - \left(\int x\,dF_\theta(x)\right)^2 \geq 0\end{aligned}$$

because the last expression is the variance of F_θ.

Second Derivative Test

Recall that the function we are trying to maximize is $a\,\theta - \kappa(\theta)$.
Its derivative is $\frac{d}{d\theta}\{a\,\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)}$.
Its second derivative is $-\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)}$.

$$\begin{aligned}\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} &= \frac{\phi''(\theta)}{\phi(\theta)} - \Big(\frac{\phi'(\theta)}{\phi(\theta)}\Big)^2 \\ &= \int x^2\,dF_\theta(x) - \Big(\int x\,dF_\theta(x)\Big)^2 \geq 0\end{aligned}$$

because the last expression is the variance of F_θ.

Assumption H2

X_1 is not a point mass at μ

$X_1 = \mu = E\,X_1$

Why we need it

In this case, F is a jump function from 0 to 1 at μ. So is F_θ for all $\theta \in (\theta_-, \theta_+)$. $\frac{d}{d\theta}\{a\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)} = a - \mu$, so either there are infinitely many critical points or none at all.

$X_1 = \mu = E\,X_1$

Why we need it

In this case, F is a jump function from 0 to 1 at μ. So is F_θ for all $\theta \in (\theta_-, \theta_+)$. $\frac{d}{d\theta}\{a\theta - \log\phi(\theta)\} = a - \frac{\phi'(\theta)}{\phi(\theta)} = a - \mu$, so either there are infinitely many critical points or none at all.

Why we can assume it

The conclusion we want is actually trivial, since $P(S_n \geq na) = 0$ for all $a > \mu$.

We can assume F is not a point mass for the interesting cases. F_θ is not a point mass either, so variance of $F_\theta > 0$.

Conclusion

We have $\frac{d^2}{d\theta^2}\{a\,\theta - \log\phi(\theta)\} = -\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} < 0.$

Conclusion

We have $\frac{d^2}{d\theta^2}\{a\,\theta - \log\phi(\theta)\} = -\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} < 0$.

$\frac{\phi'(\theta)}{\phi(\theta)}$ is strictly increasing, and $\frac{\phi'(0)}{\phi(0)} = \mu < a$, we have at most one critical point that $a = \frac{\phi'(\theta_a)}{\phi(\theta_a)}$.

Conclusion

We have $\frac{d^2}{d\theta^2}\{a\,\theta - \log\phi(\theta)\} = -\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} < 0.$

$\frac{\phi'(\theta)}{\phi(\theta)}$ is strictly increasing, and $\frac{\phi'(0)}{\phi(0)} = \mu < a$, we have at most one critical point that $a = \frac{\phi'(\theta_a)}{\phi(\theta_a)}$.

Note that the existence is not guaranteed.

Conclusion

We have $\frac{d^2}{d\theta^2}\{a\,\theta - \log\phi(\theta)\} = -\frac{d}{d\theta}\frac{\phi'(\theta)}{\phi(\theta)} < 0$.

$\frac{\phi'(\theta)}{\phi(\theta)}$ is strictly increasing, and $\frac{\phi'(0)}{\phi(0)} = \mu < a$, we have at most one critical point that $a = \frac{\phi'(\theta_a)}{\phi(\theta_a)}$.

Note that the existence is not guaranteed.

$a\,\theta - \log\phi(\theta)$ is concave, so θ_a maximize $a\,\theta - \log\phi(\theta)$, which means it gives the best bound on the rate of exponential decay.

Preview

Before discussing the existence of θ_a, We will:

Preview

Before discussing the existence of θ_a, We will:

- examine the moment generating functions $\phi(\theta)$ of some familiar distributions,

Preview

Before discussing the existence of θ_a, We will:

- examine the moment generating functions $\phi(\theta)$ of some familiar distributions,
- derive the form of $\kappa'(\theta) = \phi'(\theta)/\phi(\theta)$, which is used to optimize our upper bound on the probability of a large deviation for a particular $a > \mu$,

Preview

Before discussing the existence of θ_a, We will:

- examine the moment generating functions $\phi(\theta)$ of some familiar distributions,
- derive the form of $\kappa'(\theta) = \phi'(\theta)/\phi(\theta)$, which is used to optimize our upper bound on the probability of a large deviation for a particular $a > \mu$,
- discuss some properties of moment generating functions

Ex 1. Normal Distribution

For $X \sim \mathcal{N}(0,1)$,

$$\phi(\theta) = E\exp(\theta X) = \int e^{\theta x}(2\pi)^{-1/2}\exp(-x^2/2)dx$$
$$= \exp(\theta^2/2)\int (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)dx.$$

The integrand is the density of a normal distribution with mean θ and variance 1, so $\phi(\theta) = \exp(\theta^2/2), \theta \in (-\infty, \infty)$.

Ex 1. Normal Distribution

For $X \sim \mathcal{N}(0,1)$,

$$\begin{aligned}\phi(\theta) = E\exp(\theta X) &= \int e^{\theta x}(2\pi)^{-1/2}\exp(-x^2/2)dx \\ &= \exp(\theta^2/2)\int(2\pi)^{-1/2}\exp(-(x-\theta)^2/2)dx.\end{aligned}$$

The integrand is the density of a normal distribution with mean θ and variance 1, so $\phi(\theta) = \exp(\theta^2/2), \theta \in (-\infty, \infty)$.
Thus $\phi'(\theta)/\phi(\theta) = \theta$, and

$$F_\theta(x) = e^{-\theta^2/2}\int_{-\infty}^{x} e^{\theta y}(2\pi)^{-1/2}e^{-y^2/2}dy,$$

is a normal distribution with mean θ and variance 1.

Ex 2. Exponential Distribution with parameter λ

If $\theta < \lambda$,

$$\phi(\theta) = E\exp(\theta X) = \int_0^\infty e^{\theta x}\lambda e^{-\lambda x}dx$$
$$= \lambda/(\lambda - \theta).$$

Ex 2. Exponential Distribution with parameter λ

If $\theta < \lambda$,

$$\begin{aligned}\phi(\theta) = E\exp(\theta X) &= \int_0^\infty e^{\theta x} \lambda e^{-\lambda x} dx \\ &= \lambda/(\lambda - \theta).\end{aligned}$$

Thus $\phi'(\theta)/\phi(\theta) = 1/(\lambda - \theta)$, and

Ex 2. Exponential Distribution with parameter λ

If $\theta < \lambda$,

$$\phi(\theta) = E \exp(\theta X) = \int_0^\infty e^{\theta x} \lambda e^{-\lambda x} dx = \lambda/(\lambda - \theta).$$

Thus $\phi'(\theta)/\phi(\theta) = 1/(\lambda - \theta)$, and

$$F_\theta(x) = \frac{\lambda - \theta}{\lambda} \int_0^x e^{\theta y} \lambda e^{-\lambda y} dy = \int_0^x (\lambda - \theta) e^{-(\lambda - \theta) y} dy$$

is an exponential distribution with parameter $\lambda - \theta$.

Ex 3. Perverted Exponential

Let $g(x) = Cx^{-3}e^{-x}$ for $x \geq 1$, $g(x) = 0$ otherwise, and choose C so that g is a probability density. Then

$$\begin{aligned}\phi(\theta) = E\exp(\theta X) &= \int e^{\theta x} g(x) dx \\ &= C \int x^{-3} e^{(\theta-1)x} dx\end{aligned}$$

is finite if and only if $\theta \leq 1$.

Ex 3. Perverted Exponential

Let $g(x) = Cx^{-3}e^{-x}$ for $x \geq 1$, $g(x) = 0$ otherwise, and choose C so that g is a probability density. Then

$$\begin{aligned}\phi(\theta) = E\exp(\theta X) &= \int e^{\theta x} g(x)dx \\ &= C\int x^{-3}e^{(\theta-1)x}dx\end{aligned}$$

is finite if and only if $\theta \leq 1$.
When $\theta \leq 1$,

$$\frac{\phi'(\theta)}{\phi(\theta)} \leq \frac{\phi'(1)}{\phi(1)} = \int_1^\infty Cx^{-2}dx \Big/ \int_1^\infty Cx^{-3}dx = 2.$$

Properties of Moment Generating Functions

Let $x_0 = \sup\{x : F(x) < 1\}$. If $x_0 < \infty$, then:

- $\phi(\theta) < \infty$ for all $\theta > 0$,
- $\phi'(\theta)/\phi(\theta) \to x_0$ as $\theta \uparrow \infty$.

Properties of Moment Generating Functions

Let $x_0 = \sup\{x : F(x) < 1\}$. If $x_0 < \infty$, then:

- $\phi(\theta) < \infty$ for all $\theta > 0$,
- $\phi'(\theta)/\phi(\theta) \to x_0$ as $\theta \uparrow \infty$.

Outline of proof

If $x_0 < \infty$, then $P(X > x_0) = 0$.
Then $\phi(\theta) = \int e^{\theta x} dF(x) = \int_{-\infty}^{x_0} e^{\theta x} dF(x) < \infty$ for all θ.

Properties of Moment Generating Functions

Let $x_0 = \sup\{x : F(x) < 1\}$. If $x_0 < \infty$, then:

- $\phi(\theta) < \infty$ for all $\theta > 0$,
- $\phi'(\theta)/\phi(\theta) \to x_0$ as $\theta \uparrow \infty$.

Outline of proof

If $x_0 < \infty$, then $P(X > x_0) = 0$.
Then $\phi(\theta) = \int e^{\theta x} dF(x) = \int_{-\infty}^{x_0} e^{\theta x} dF(x) < \infty$ for all θ.

Furthermore,

$$\frac{\phi'(\theta)}{\phi(\theta)} = \frac{\int_{-\infty}^{x_0} x e^{\theta x} dF(x)}{\int_{-\infty}^{x_0} e^{\theta x} dF(x)},$$

with F putting nonzero weight near x_0. As $\theta \to \infty$, the tail is growing faster than the rest, with the numerator scaled by x_0.

Interpretations

For the normal and exponential distributions,
$\sup\{x : F(x) < 1\} = \infty$.

- Thus we have $\phi'(\theta)/\phi(\theta) \to \infty$ as $\theta \to \theta_+$, and
- we can solve $a = \phi'(\theta)/\phi(\theta)$ for any $a > \mu$.

Interpretations

For the normal and exponential distributions, $\sup\{x : F(x) < 1\} = \infty$.

- Thus we have $\phi'(\theta)/\phi(\theta) \to \infty$ as $\theta \to \theta_+$, and
- we can solve $a = \phi'(\theta)/\phi(\theta)$ for any $a > \mu$.

In Example 3, we cannot solve $a = \phi'(\theta)/\phi(\theta)$ for $a > 2$

Interpretations

For the normal and exponential distributions, $\sup\{x : F(x) < 1\} = \infty$.

- Thus we have $\phi'(\theta)/\phi(\theta) \to \infty$ as $\theta \to \theta_+$, and
- we can solve $a = \phi'(\theta)/\phi(\theta)$ for any $a > \mu$.

In Example 3, we cannot solve $a = \phi'(\theta)/\phi(\theta)$ for $a > 2$

(H3) there is a $\theta_a \in (0, \theta_+)$ so that $a = \phi'(\theta_a)/\phi(\theta_a)$.

Optimal exponential decay

Optimal exponential decay

Theorem

Suppose that (H1), (H2) and (H3) hold. Then,

$$\gamma(a) = \lim_{n\to\infty} n^{-1} \log P(S_n \geq na) = -a\theta_a + \log \phi(\theta_a).$$

Optimal exponential decay

Theorem

Suppose that (H1), (H2) and (H3) hold. Then,

$$\gamma(a) = \lim_{n\to\infty} n^{-1} \log P(S_n \geq na) = -a\theta_a + \log \phi(\theta_a).$$

Informally, the theorem states that if $a = \phi'(\theta_a)/\phi(\theta_a)$, then (asymptotically) the probability of a large deviation decays as exponentially fast as possible.

Proof of Theorem

Proof of Theorem

(lim sup). Earlier using Chebyshev's inequality we observed that $P(S_n \geq na) \leq \exp(-n\{a\theta - \log\phi(\theta)\})$, for $\theta \in (0, \theta_+)$, and in particular for θ_a. This implies

$$\limsup_{n\to\infty} n^{-1} \log P(S_n \geq na) \leq -a\theta_a + \log\phi(\theta_a).$$

Proof of Theorem

(lim sup). Earlier using Chebyshev's inequality we observed that $P(S_n \geq na) \leq \exp(-n\{a\theta - \log\phi(\theta)\})$, for $\theta \in (0, \theta_+)$, and in particular for θ_a. This implies

$$\limsup_{n\to\infty} n^{-1} \log P(S_n \geq na) \leq -a\theta_a + \log\phi(\theta_a).$$

(lim inf). The other direction requires a bit more work. Fix $\lambda \in (\theta_a, \theta_+)$ and let $X_1^\lambda, X_2^\lambda, \ldots$ be i.i.d. with distribution F_λ (well-defined by H1); set $S_n^\lambda = X_1^\lambda + \cdots + X_n^\lambda$.

Proof of Theorem

(lim sup). Earlier using Chebyshev's inequality we observed that $P(S_n \geq na) \leq \exp(-n\{a\theta - \log\phi(\theta)\})$, for $\theta \in (0, \theta_+)$, and in particular for θ_a. This implies

$$\limsup_{n\to\infty} n^{-1} \log P(S_n \geq na) \leq -a\theta_a + \log\phi(\theta_a).$$

(lim inf). The other direction requires a bit more work. Fix $\lambda \in (\theta_a, \theta_+)$ and let $X_1^\lambda, X_2^\lambda, \ldots$ be i.i.d. with distribution F_λ (well-defined by H1); set $S_n^\lambda = X_1^\lambda + \cdots + X_n^\lambda$. Before we proceed, a short digression on...

Abs. Cts. Measures & Radon-Nikodym Derivative

Abs. Cts. Measures & Radon-Nikodym Derivative

Let $(X, \mathcal{F})$ be a measure space equipped with two measures, μ and ν. By definition, we say that ν is *absolutely continuous with respect to* μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{F}$, and we write $\nu \ll \mu$.

Abs. Cts. Measures & Radon-Nikodym Derivative

Let $(X, \mathcal{F})$ be a measure space equipped with two measures, μ and ν. By definition, we say that ν is *absolutely continuous with respect to* μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{F}$, and we write $\nu \ll \mu$. If we simply say μ is *absolutely continuous* then we mean with respect to the Lebesgue measure.

Example

- Any measure is (trivially) absolutely continuous with respect to itself.

Abs. Cts. Measures & Radon-Nikodym Derivative

Let $(X, \mathcal{F})$ be a measure space equipped with two measures, μ and ν. By definition, we say that ν is *absolutely continuous with respect to* μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{F}$, and we write $\nu \ll \mu$. If we simply say μ is *absolutely continuous* then we mean with respect to the Lebesgue measure.

Example

- Any measure is (trivially) absolutely continuous with respect to itself.
- A finite measure μ is absolutely continuous iff the function $F(x) = \mu((-\infty, x]))$ is absolutely continuous as a function.

Abs. Cts. Measures & Radon-Nikodym Derivative

Let $(X, \mathcal{F})$ be a measure space equipped with two measures, μ and ν. By definition, we say that ν is *absolutely continuous with respect to* μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{F}$, and we write $\nu \ll \mu$. If we simply say μ is *absolutely continuous* then we mean with respect to the Lebesgue measure.

Example

- Any measure is (trivially) absolutely continuous with respect to itself.
- A finite measure μ is absolutely continuous iff the function $F(x) = \mu((-\infty, x]))$ is absolutely continuous as a function.
- Hence the Gaussian measure is absolutely continuous, but the "Devil's staircase" measure is not (because it assigns positive measure to the Cantor set).

Abs. Cts. Measures & Radon-Nikodym Derivative

Abs. Cts. Measures & Radon-Nikodym Derivative

Key result (Radon-Nikodym, 1930): if ν and μ are σ-finite, and ν is absolutely continuous w.r.t μ, then there exists a measurable function $f : X \to [0, \infty)$ such that for measurable subsets $A \subset X$,

$$\nu(A) = \int_A f d\mu$$

Abs. Cts. Measures & Radon-Nikodym Derivative

Key result (Radon-Nikodym, 1930): if ν and μ are σ-finite, and ν is absolutely continuous w.r.t μ, then there exists a measurable function $f : X \to [0, \infty)$ such that for measurable subsets $A \subset X$,

$$\nu(A) = \int_A f d\mu$$

The function f is called the *Radon-Nikodym derivative* and we write $f = d\nu/d\mu$. It represents a sort of "rate of change of measure", which is why it's called a derivative.

Example

- Application to probability: Any distribution that admits a density is absolutely continuous, and the Radon-Nikodym derivative is the density function.

Application to Current Theorem

Application to Current Theorem

Recall that,

$$F_\lambda(x) = \frac{1}{\phi(\lambda)} \int_{-\infty}^{x} e^{\lambda y} dF(y).$$

It follows that $dF_\lambda / dF = e^{\lambda x} / \phi(\lambda)$.

Application to Current Theorem

Recall that,

$$F_\lambda(x) = \frac{1}{\phi(\lambda)} \int_{-\infty}^{x} e^{\lambda y}\, dF(y).$$

It follows that $dF_\lambda/dF = e^{\lambda x}/\phi(\lambda)$. In our case, since $dF \ll dF_\lambda$ and $dF_\lambda \ll dF$, we may write $dF/dF_\lambda = e^{-\lambda x}\phi(\lambda)$.

Application to Current Theorem

Recall that,

$$F_\lambda(x) = \frac{1}{\phi(\lambda)} \int_{-\infty}^{x} e^{\lambda y} dF(y).$$

It follows that $dF_\lambda / dF = e^{\lambda x}/\phi(\lambda)$. In our case, since $dF \ll dF_\lambda$ and $dF_\lambda \ll dF$, we may write $dF/dF_\lambda = e^{-\lambda x}\phi(\lambda)$. Let F_λ^n and F^n denote the distributions of S_n^λ and S_n respectively. We claim that the following is true:

$$\frac{dF_\lambda^n}{dF^n}(x) = e^{-\lambda x}\phi(\lambda)^n.$$

Application to Current Theorem

Recall that,

$$F_\lambda(x) = \frac{1}{\phi(\lambda)} \int_{-\infty}^{x} e^{\lambda y} dF(y).$$

It follows that $dF_\lambda / dF = e^{\lambda x}/\phi(\lambda)$. In our case, since $dF \ll dF_\lambda$ and $dF_\lambda \ll dF$, we may write $dF/dF_\lambda = e^{-\lambda x}\phi(\lambda)$. Let F_λ^n and F^n denote the distributions of S_n^λ and S_n respectively. We claim that the following is true:

$$\frac{dF_\lambda^n}{dF^n}(x) = e^{-\lambda x}\phi(\lambda)^n.$$

The above arguments show that this holds when $n = 1$ so naturally...

Proof of Lemma

We will induct on n.

Proof of Lemma

We will induct on n.

$$F^n(z) = F^{n-1} * F(z) = \int_{-\infty}^{\infty} F(z-x)dF^{n-1}(x)$$

□

Proof of Lemma

We will induct on n.

$$F^n(z) = F^{n-1} * F(z) = \int_{-\infty}^{\infty} F(z-x)dF^{n-1}(x)$$
$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} dF(y) \right) dF^{n-1}(x)$$

□

Proof of Lemma

We will induct on n.

$$\begin{aligned} F^n(z) = F^{n-1} * F(z) &= \int_{-\infty}^{\infty} F(z-x)dF^{n-1}(x) \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} dF(y) \right) dF^{n-1}(x) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} e^{-\lambda(x+y)} \phi(\lambda)^n dF_\lambda(y) dF_\lambda^{n-1}(x) \end{aligned}$$

□

Proof of Lemma

We will induct on n.

$$
\begin{aligned}
F^n(z) = F^{n-1} * F(z) &= \int_{-\infty}^{\infty} F(z-x)dF^{n-1}(x) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} dF(y) \right) dF^{n-1}(x) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} e^{-\lambda(x+y)} \phi(\lambda)^n dF_\lambda(y) dF_\lambda^{n-1}(x) \\
&= E\left(1_{\{S_{n-1}^\lambda + X_n^\lambda \le z\}} e^{-\lambda(S_{n-1}^\lambda + X_n^\lambda)} \phi(\lambda)^n \right)
\end{aligned}
$$

□

Proof of Lemma

We will induct on n.

$$
\begin{aligned}
F^n(z) = F^{n-1} * F(z) &= \int_{-\infty}^{\infty} F(z-x)\,dF^{n-1}(x) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} dF(y) \right) dF^{n-1}(x) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} e^{-\lambda(x+y)} \phi(\lambda)^n dF_\lambda(y) dF_\lambda^{n-1}(x) \\
&= E\left(1_{\{S_{n-1}^\lambda + X_n^\lambda \le z\}} e^{-\lambda(S_{n-1}^\lambda + X_n^\lambda)} \phi(\lambda)^n \right) \\
&= \int_{-\infty}^{z} e^{-\lambda u} \phi(\lambda)^n dF_\lambda^n(u)
\end{aligned}
$$

□

Proof of Theorem

Proof of Theorem

Let $b > a$. Using the lemma, we observe that

$$P(S_n \geq na) \geq \int_{na}^{nb} e^{-\lambda x}\phi(\lambda)^n dF_\lambda^n(x) \geq \phi(\lambda)^n e^{-\lambda nb}(F_\lambda^n(nb) - F_\lambda^n(na)).$$

Proof of Theorem

Let $b > a$. Using the lemma, we observe that

$$P(S_n \geq na) \geq \int_{na}^{nb} e^{-\lambda x} \phi(\lambda)^n dF_\lambda^n(x) \geq \phi(\lambda)^n e^{-\lambda nb}(F_\lambda^n(nb) - F_\lambda^n(na)).$$

By construction F_λ has mean $\phi(\lambda)'/\phi(\lambda)$. So we choose b such that $a < \phi(\lambda)'/\phi(\lambda) < b$. Then, by the SLLN, it follows that $F_\lambda^n(nb) - F_\lambda^n(na) \to 1$ as $n \to \infty$.

Proof of Theorem

Let $b > a$. Using the lemma, we observe that

$$P(S_n \geq na) \geq \int_{na}^{nb} e^{-\lambda x}\phi(\lambda)^n dF_\lambda^n(x) \geq \phi(\lambda)^n e^{-\lambda nb}(F_\lambda^n(nb) - F_\lambda^n(na)).$$

By construction F_λ has mean $\phi(\lambda)'/\phi(\lambda)$. So we choose b such that $a < \phi(\lambda)'/\phi(\lambda) < b$. Then, by the SLLN, it follows that $F_\lambda^n(nb) - F_\lambda^n(na) \to 1$ as $n \to \infty$. Therefore,

$$\liminf_{n\to\infty} n^{-1} \log(P(S_n \geq na)) \geq -\lambda b + \log\phi(\lambda).$$

Proof of Theorem

Let $b > a$. Using the lemma, we observe that

$$P(S_n \geq na) \geq \int_{na}^{nb} e^{-\lambda x}\phi(\lambda)^n dF_\lambda^n(x) \geq \phi(\lambda)^n e^{-\lambda nb}(F_\lambda^n(nb) - F_\lambda^n(na)).$$

By construction F_λ has mean $\phi(\lambda)'/\phi(\lambda)$. So we choose b such that $a < \phi(\lambda)'/\phi(\lambda) < b$. Then, by the SLLN, it follows that $F_\lambda^n(nb) - F_\lambda^n(na) \to 1$ as $n \to \infty$. Therefore,

$$\liminf_{n\to\infty} n^{-1} \log(P(S_n \geq na)) \geq -\lambda b + \log\phi(\lambda).$$

Now take $\lambda > \theta_a$ to be arbitrary, and then $b > a$ to be arbitrary to finish the proof. □

Sans H3

Sans H3

In the case when H3 cannot be assumed we suppose the following:

Assumption

(H4). If $a\theta - \log\phi(\theta)$ cannot be maximized then assume $x_0 = \infty$, $\theta_+ < \infty$ and $\phi'(\theta)/\phi(\theta) \uparrow a_0 < \infty$ as $\theta \uparrow \theta_+$.

Sans H3

In the case when H3 cannot be assumed we suppose the following:

Assumption

(H4). If $a\theta - \log\phi(\theta)$ cannot be maximized then assume $x_0 = \infty$, $\theta_+ < \infty$ and $\phi'(\theta)/\phi(\theta) \uparrow a_0 < \infty$ as $\theta \uparrow \theta_+$.

Theorem

Assuming H1, H2 and H4, if $a_0 \leq a < \infty$ then, $\gamma(a) = -a\theta_+ + \log\phi(\theta_+)$. *That is, $\gamma(a)$ is linear for $a \geq a_0$.*

Sans H3

In the case when H3 cannot be assumed we suppose the following:

Assumption

(H4). If $a\theta - \log\phi(\theta)$ cannot be maximized then assume $x_0 = \infty$, $\theta_+ < \infty$ and $\phi'(\theta)/\phi(\theta) \uparrow a_0 < \infty$ as $\theta \uparrow \theta_+$.

Theorem

Assuming H1, H2 and H4, if $a_0 \leq a < \infty$ then, $\gamma(a) = -a\theta_+ + \log\phi(\theta_+)$. *That is, $\gamma(a)$ is linear for $a \geq a_0$.*

It can be shown that if $EX_1 = 0$ and $\phi(\theta) = \infty$ for all $\theta > 0$ then $n^{-1}\log P(S_n \geq na) \to 0$ for all $a > 0$.

Sans H3

In the case when H3 cannot be assumed we suppose the following:

Assumption

(H4). If $a\theta - \log\phi(\theta)$ cannot be maximized then assume $x_0 = \infty$, $\theta_+ < \infty$ and $\phi'(\theta)/\phi(\theta) \uparrow a_0 < \infty$ as $\theta \uparrow \theta_+$.

Theorem

Assuming H1, H2 and H4, if $a_0 \leq a < \infty$ then, $\gamma(a) = -a\theta_+ + \log\phi(\theta_+)$. *That is, $\gamma(a)$ is linear for $a \geq a_0$.*

It can be shown that if $EX_1 = 0$ and $\phi(\theta) = \infty$ for all $\theta > 0$ then $n^{-1}\log P(S_n \geq na) \to 0$ for all $a > 0$. This shows that H1 is the correct assumption to make.

Recap

To get a feel for what the answers look like, we revisit our examples. Recall the notation

$$\kappa(\theta) = \log \phi(\theta) \qquad \kappa'(\theta) = \phi'(\theta)/\phi(\theta) \qquad \theta_a \text{ solves } \kappa'(\theta_a) = a$$

$$\gamma(a) = \lim_{n\to\infty} (1/n) \log P(S_n \geq na) = -a\theta_a + \kappa(\theta_a),$$

that is,

$$P(S_n \geq na) \leq e^{n\gamma(a)}.$$

Normal distribution

$\phi(\theta) = \exp(\theta^2/2)$, so

$$\kappa(\theta) = \theta^2/2 \qquad \kappa'(\theta) = \theta \qquad \theta_a = a$$

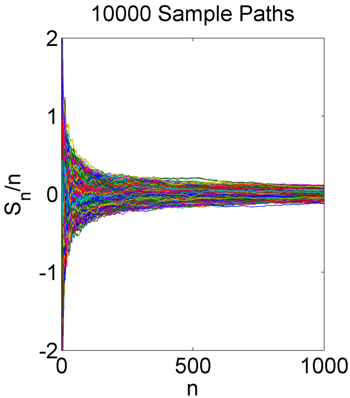
$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a^2/2.$$

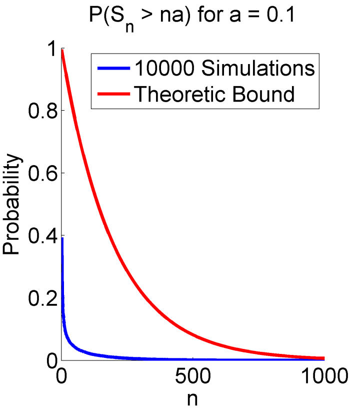
Normal distribution

$\phi(\theta) = \exp(\theta^2/2)$, so

$$\kappa(\theta) = \theta^2/2 \qquad \kappa'(\theta) = \theta \qquad \theta_a = a$$

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a^2/2.$$



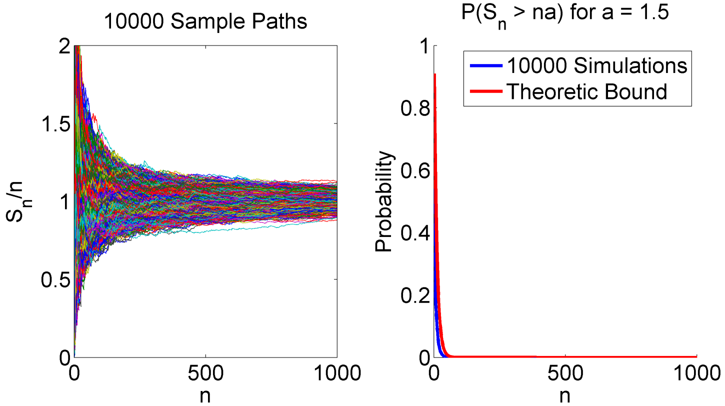


Exponential distribution with $\lambda = 1$

$\phi(\theta) = \lambda/(\lambda - \theta)$, so

$$\kappa(\theta) = -\log(1-\theta) \qquad \kappa'(\theta) = 1/(1-\theta) \qquad \theta_a = 1 - 1/a$$

$$\gamma(a) = -a\theta_a + \kappa(\theta_a) = -a + 1 + \log a.$$



References

Richard Durrett.
"Large Deviations." *Probability Theory and Examples.*
Cambridge: Cambridge UP, 2010. 86-93. Print.