Some Basic Statistical Theorems

Review

Review

Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.

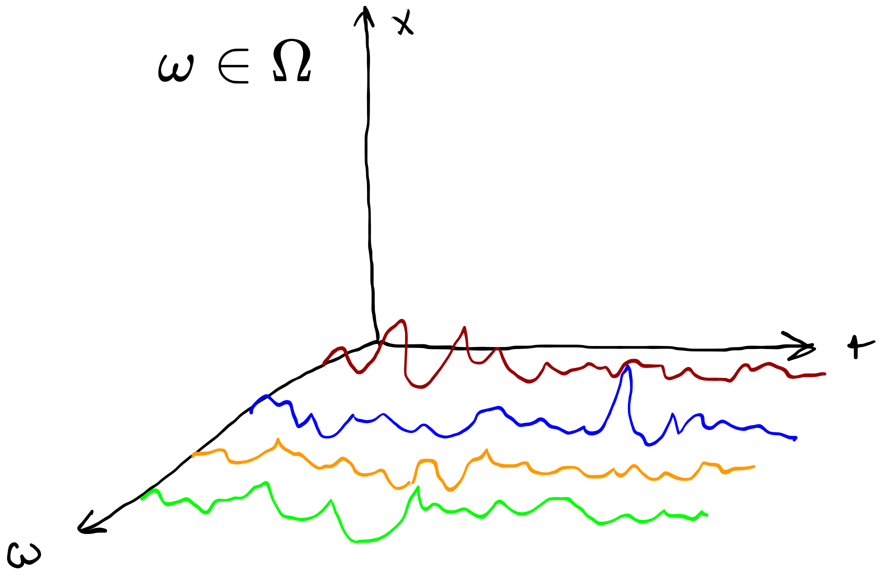
Review

Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.

$\omega \in \Omega$

Review

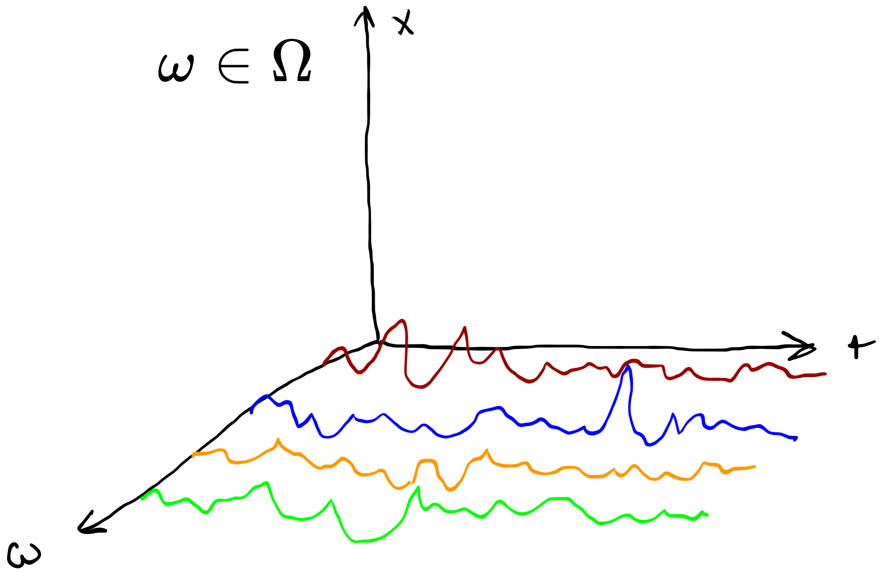
Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.



Review

Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.

$\omega \in \Omega$



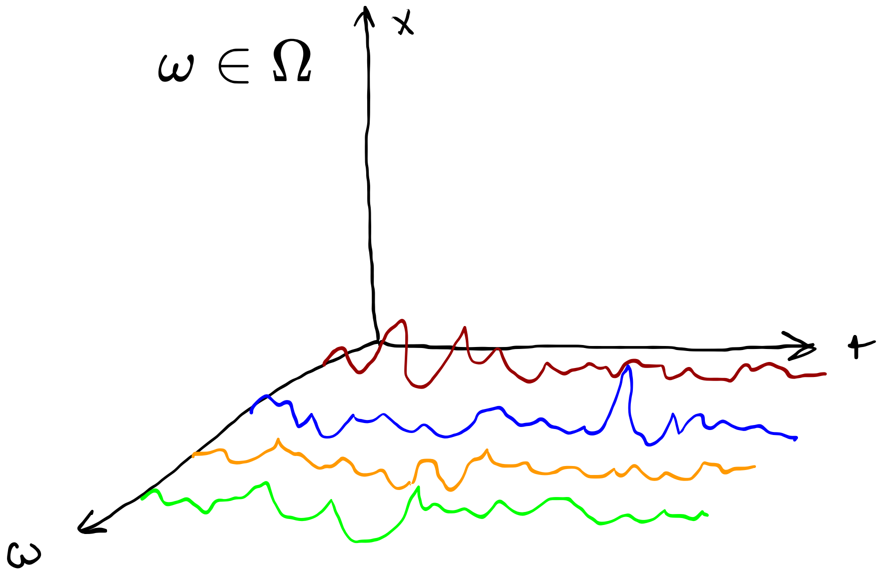
$\mathcal{F}$

$\vdots$

Review

Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.

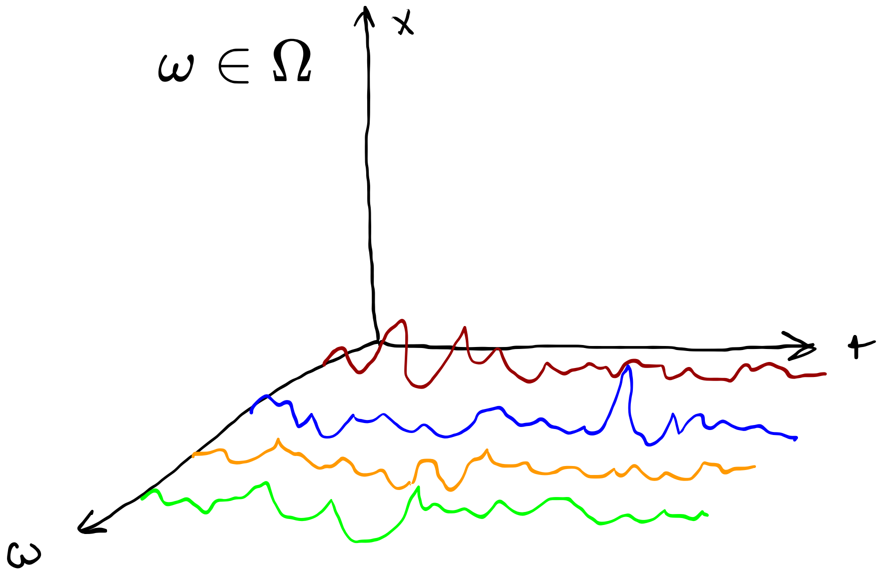
$\omega \in \Omega$

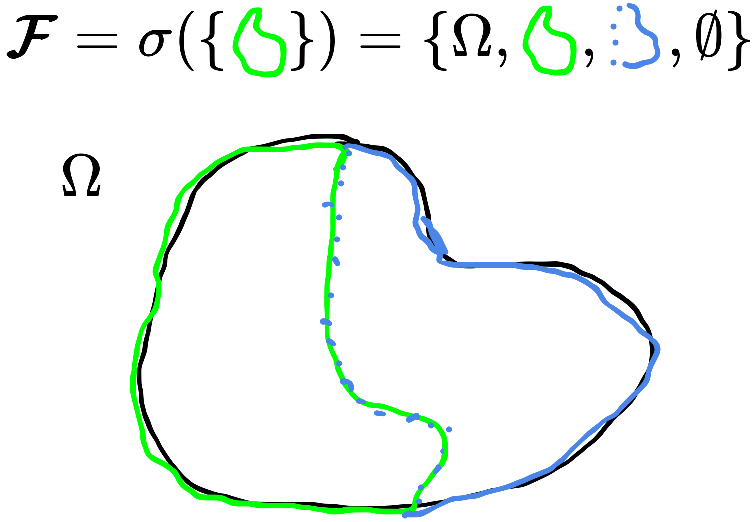


$$\mathcal{F} = \sigma(\{\ \}) = \{\Omega, \ , \ , \emptyset\}$$

Review

Given a **probability space** $(\Omega, \mathcal{F}, P)$, and a **measurable space** $(E, \mathcal{E})$, an E-valued **random variable is a measurable function** $X : \Omega \to E$.





Review

Convergence:

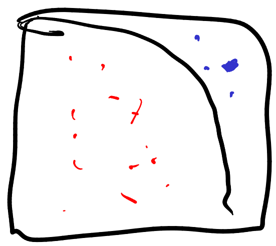
- Sure ("pointwise")
- Almost Sure
- In Probability
- Weak ("in distribution"/"in law")

Review

Convergence:

- Sure ("pointwise") $X_n(\omega) \to X(\omega) \quad \forall\, \omega \in \Omega$
- Almost Sure
- In Probability
- Weak ("in distribution"/"in law")

Review



Convergence:

- Sure ("pointwise") $X_n(\omega) \to X(\omega) \quad \forall\, \omega \in \Omega$
- Almost Sure $P(\{\omega : X_n(\omega) \to X\}) = 1$
- In Probability
- Weak ("in distribution"/"in law")

Review

Convergence:

- Sure ("pointwise") $X_n(\omega) \to X(\omega) \quad \forall\, \omega \in \Omega$
- Almost Sure $P(\{\omega : X_n(\omega) \to X\}) = 1$
- In Probability $P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \to 0 \quad \forall \epsilon > 0$
- Weak ("in distribution"/"in law")

Review

Convergence:

- Sure ("pointwise") $X_n(\omega) \to X(\omega) \quad \forall\, \omega \in \Omega$
- Almost Sure $P(\{\omega : X_n(\omega) \to X\}) = 1$
- In Probability $P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \to 0 \quad \forall \epsilon > 0$
- Weak ("in distribution"/"in law")

$F_{X_n}(\alpha) \to F_X(\alpha)$ for each continuity point.

Outline for Today

Outline for Today

1. Concentration Inequalities

Outline for Today

1. Concentration Inequalities
2. Proof of the weak law of large numbers and $\frac{1}{\sqrt{N}}$ convergence rates

Outline for Today

1. Concentration Inequalities

Break

2. Proof of the weak law of large numbers and $\frac{1}{\sqrt{N}}$ convergence rates

Outline for Today

1. Concentration Inequalities

Break

2. Proof of the weak law of large numbers and $\frac{1}{\sqrt{N}}$ convergence rates
3. Central limit theorem
4. Importance sampling

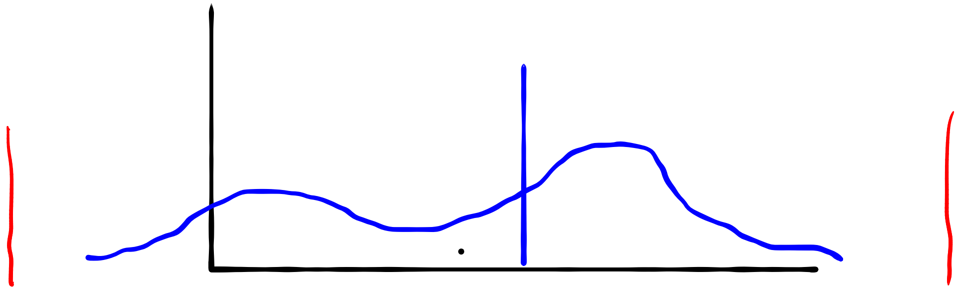
Concentration Inequalities

Concentration Inequalities

Intuition: If an r.v. has a finite variance, the probability that a random variable takes a value far from its mean should be small

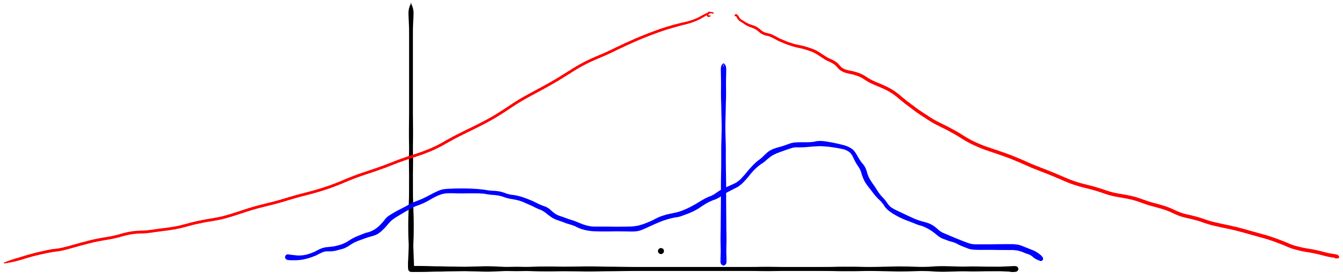
Concentration Inequalities

Intuition: If an r.v. has a finite variance, the probability that a random variable takes a value far from its mean should be small



Concentration Inequalities

Intuition: If an r.v. has a finite variance, the probability that a random variable takes a value far from its mean should be small



Concentration inequalities take the form

$$P(X \geq t) \leq \phi(t)$$

where ϕ goes to zero (quickly) as $t \to \infty$

Concentration Inequalities

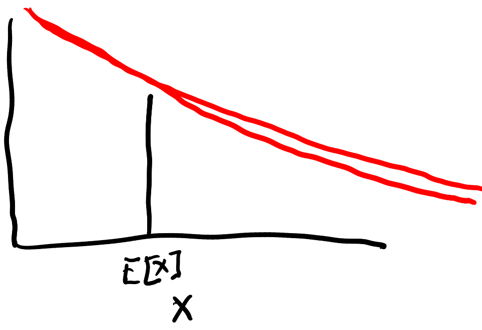
Concentration Inequalities

Markov's Inequality:

If $X \geq 0$, then

$$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall t \geq 0$$

X=1



Proof:

$$E[X] = \int_{-\infty}^{\infty} x\,p(x)\,dx$$

$$= \int_0^{\infty} x\,p(x)\,dx$$

$$= \underbrace{\int_0^{a} x\,p(x)\,dx}_{\geq 0} + \int_a^{\infty} x\,p(x)\,dx$$

$$\geq \int_a^{\infty} x\,p(x)\,dx \geq \int_a^{\infty} a\,p(x)\,dx = a\int_a^{\infty} p(x)\,dx$$

$$= a\,P(X \geq a)$$

$$\therefore \frac{E[X]}{a} \geq P(X \geq a)$$

Concentration Inequalities

Concentration Inequalities

Chebyshev's Inequality:

Let X be *any* real-valued random variable with $\mathrm{Var}(X) < \infty$. Then

Concentration Inequalities

Chebyshev's Inequality:

Let X be *any* real-valued random variable with $\mathrm{Var}(X) < \infty$. Then

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

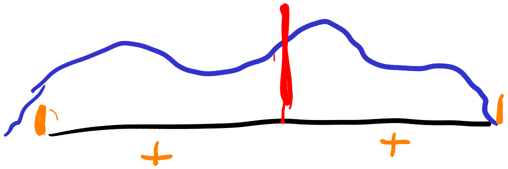
Proof: Recall $\mathrm{Var}(X) = E[(X - E[X])^2]$

Use Markov's inequality on $(X-E[X])^2$

$$P\big(\overset{Y}{(X-E[X])^2} \geq \overset{\gamma}{t^2}\big) \leq \frac{E[(X-E[X])^2]}{t^2} \qquad \frac{E[Y]}{\gamma}$$

$$P(|X-E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$$

$$0.05 = \frac{\mathrm{Var}(X)}{t^2}$$



Concentration Inequalities

Chebyshev's Inequality:

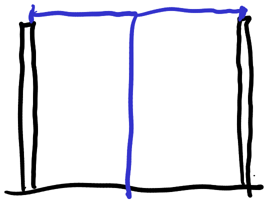
Let X be *any* real-valued random variable with $\mathrm{Var}(X) < \infty$. Then

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Very general, but still loose

Concentration Inequalities

k	Min. % within *k* standard deviations of mean	Max. % beyond *k* standard deviations from mean
1	0%	100%
$\sqrt{2}$	50%	50%
1.5	55.56%	44.44%
2	75%	25%
$2\sqrt{2}$	87.5%	12.5%
3	88.8889%	11.1111%
4	93.75%	6.25%
5	96%	4%
6	97.2222%	2.7778%
7	97.9592%	2.0408%
8	98.4375%	1.5625%
9	98.7654%	1.2346%
10	99%	1%



Concentration Inequalities

k	Min. % within *k* standard deviations of mean	Max. % beyond *k* standard deviations from mean
1	0%	100%
$\sqrt{2}$	50%	50%
1.5	55.56%	44.44%
2	75%	25%
$2\sqrt{2}$	87.5%	12.5%
3	88.8889%	11.1111%
4	93.75%	6.25%
5	96%	4%
6	97.2222%	2.7778%
7	97.9592%	2.0408%
8	98.4375%	1.5625%
9	98.7654%	1.2346%
10	99%	1%

For Normal Distribution

Concentration Inequalities

k	Min. % within *k* standard deviations of mean	Max. % beyond *k* standard deviations from mean
1	0%	100%
$\sqrt{2}$	50%	50%
1.5	55.56%	44.44%
2	75%	25%
$2\sqrt{2}$	87.5%	12.5%
3	88.8889%	11.1111%
4	93.75%	6.25%
5	96%	4%
6	97.2222%	2.7778%
7	97.9592%	2.0408%
8	98.4375%	1.5625%
9	98.7654%	1.2346%
10	99%	1%

68% For Normal Distribution

95%

99%

Concentration Inequalities

Concentration Inequalities

Moment generating function: $M_X(t) \equiv E[e^{tX}]$

Concentration Inequalities

Moment generating function: $M_X(t) \equiv E[e^{tX}]$

Chernoff Bound: If the moment-generating function M_X exists, then

$$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t > 0$$

$\forall a$

infinite number of bounds, one for each t

and

$$P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t < 0$$

Proof: Apply Markov's inequality to e^{tX}

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}}$$

Concentration Inequalities

Moment generating function: $M_X(t) \equiv E[e^{tX}]$

Chernoff Bound: If the moment-generating function M_X exists, then

$$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t > 0$$

and

$$P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t < 0$$

Tighter than Markov and Chebyshev

Concentration Inequalities

Concentration Inequalities

Name	Requirements	Bound

Concentration Inequalities

Name	Requirements	Bound
Markov	$X \geq 0$, $\mathrm{E}[X]$ exists	$P(X \geq t) \leq \dfrac{E[X]}{t} \quad \forall\, t \geq 0$

Concentration Inequalities

Name	Requirements	Bound
Markov	$X \geq 0, \ \mathrm{E}[X]$ exists	$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall\, t \geq 0$
Chebyshev	$\mathrm{Var}(X) < \infty$	$P(\lvert X - E[X] \rvert \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$

Concentration Inequalities

Name	Requirements	Bound
Markov	$X \geq 0$, $\mathrm{E}[X]$ exists	$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall\, t \geq 0$
Chebyshev	$\mathrm{Var}(X) < \infty$	$P(\lvert X - E[X] \rvert \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$
Chernoff	M_X exists	$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t > 0$ $P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t < 0$

Concentration Inequalities

Name	Requirements	Bound
Markov	$X \geq 0,\ \mathrm{E}[X]$ exists	$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall\, t \geq 0$
Chebyshev	$\mathrm{Var}(X) < \infty$	$P(\lvert X - E[X]\rvert \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$
Chernoff	M_X exists	$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t > 0$ $P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t < 0$

More restrictive

Concentration Inequalities

Name	Requirements	Bound
Markov	$X \geq 0,\ \mathrm{E}[X]$ exists	$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall\, t \geq 0$
Chebyshev	$\mathrm{Var}(X) < \infty$	$P(\lvert X - E[X]\rvert \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$
Chernoff	M_X exists	$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t > 0$ $P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall\, t < 0$

More restrictive

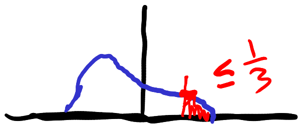
Tighter

Break

Name	Requirements	Bound
Markov	$X \geq 0$, $\mathrm{E}[X]$ exists	$P(X \geq t) \leq \frac{E[X]}{t} \quad \forall t \geq 0$
Chebyshev	$\mathrm{Var}(X) < \infty$	$P(\lvert X - E[X]\rvert \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$
Chernoff	M_X exists	$P(X \geq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall t > 0$ $P(X \leq a) \leq \frac{E[e^{tX}]}{e^{ta}} \quad \forall t < 0$

Let Y be a r.v. that takes values in $[-1, 1]$ with mean -0.5. Give an upper bound on the probability that $Y \geq 0.5$.

Let $Z = Y+1$ $\quad P(Y \geq 0.5) = P(Z \geq 1.5) \leq \frac{0.5}{1.5} = \frac{1}{3}$



(Weak) Law of large numbers

(Weak) Law of large numbers

Let X_i be independent identically distributed r.v.s with mean μ and variance σ^2. If $Q_N \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$, then $Q_N \to_p \mu$.

(Weak) Law of large numbers

Let X_i be independent identically distributed r.v.s with mean μ and variance σ^2. If $Q_N \equiv \frac{1}{N}\sum_{i=1}^{N} X_i$, then $Q_N \to_p \mu$.

Proof:

$$\begin{aligned}
\mathrm{Var}(Q_N) &= \mathrm{Var}\left(\frac{1}{N}\sum_i X_i\right) \\
&= \frac{1}{N^2}\mathrm{Var}\left(\sum_i X_i\right) \\
&= \frac{1}{N^2}\sum_i \mathrm{Var}(X_i) \quad \text{(Bienaymé)} \\
&= \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N}
\end{aligned}$$

$E[Q_N] = \mu$

$$\begin{aligned}
P(|Q_N - \mu| \geq \varepsilon) &\leq \frac{\mathrm{Var}(Q_N)}{\varepsilon^2} \quad \text{(Chebyshev)} \\
&= \frac{\sigma^2}{N\varepsilon^2}
\end{aligned}$$

$P(|Q_N - \mu| \geq \varepsilon) \to 0$ as $N \to \infty$ $\forall \varepsilon > 0$ $\therefore Q_N \to_p \mu$ $\square$

Law of Large Numbers

Two somewhat astounding takeaways:

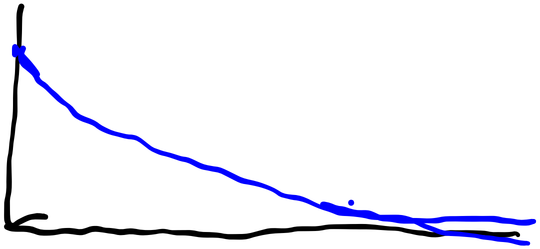
Law of Large Numbers

Two somewhat astounding takeaways:

Law of Large Numbers

Two somewhat astounding takeaways:

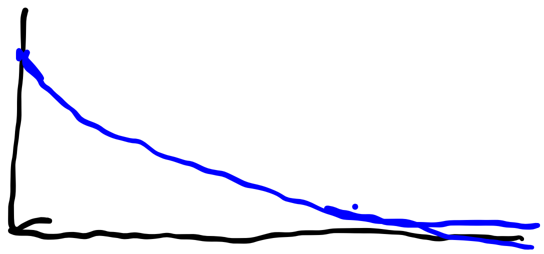
1. Standard deviation decays at $\frac{1}{\sqrt{N}}$ *regardless of dimension*.



Law of Large Numbers

Two somewhat astounding takeaways:

1. Standard deviation decays at $\frac{1}{\sqrt{N}}$ *regardless of dimension*.



2. You can estimate the "standard error" with

$$SE = \frac{s}{\sqrt{N}}$$

empirical estimate of standard deviaton of X_i

where s is the sample standard deviation.

Confidence Intervals

Confidence Intervals

How do you estimate $|Q_N - \mu|$?

Confidence Intervals

How do you estimate $|Q_N - \mu|$?

Given a random variable Q, a γ **Confidence Interval**, $[u(Q), v(Q)]$, is a random interval that contains μ with probability γ, i.e.

$$P(u(Q) \leq \mu \leq v(Q)) = \gamma$$

Confidence Intervals

How do you estimate $|Q_N - \mu|$?

Given a random variable Q, a γ **Confidence Interval**, $[u(Q), v(Q)]$, is a random interval that contains μ with probability γ, i.e.

$$P(u(Q) \leq \mu \leq v(Q)) = \gamma$$

Example: $Q_N \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$

Confidence Intervals

How do you estimate $|Q_N - \mu|$?

Given a random variable Q, a γ **Confidence Interval**, $[u(Q), v(Q)]$, is a random interval that contains μ with probability γ, i.e.

$$P(u(Q) \leq \mu \leq v(Q)) = \gamma$$

Example: $Q_N \equiv \frac{1}{N} \sum_{i=1}^{N} X_i$

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} = 1 - \gamma = 0.05$$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} = 1 - \gamma = 0.05$$

$$t = \sqrt{\frac{\mathrm{Var}(X)}{0.05}}$$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} = 1 - \gamma = 0.05$$

$$t = \sqrt{\frac{\mathrm{Var}(X)}{0.05}}$$

$$t \approx \frac{SE}{\sqrt{0.05}} \approx 4.47 SE$$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} = 1 - \gamma = 0.05$$

$$t = \sqrt{\frac{\mathrm{Var}(X)}{0.05}}$$

$$t \approx \frac{SE}{\sqrt{0.05}} \approx 4.47SE$$

Approximate 95% CI: $[Q_N - 4.47\,SE, Q_N + 4.47\,SE]$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use Chebyshev.

Use $\gamma = 0.95$

$$P(|X - E[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2} = 1 - \gamma = 0.05$$

$$t = \sqrt{\frac{\mathrm{Var}(X)}{0.05}}$$

$$t \approx \frac{SE}{\sqrt{0.05}} \approx 4.47 SE$$

Approximate 95% CI: $[Q_N - 4.47\,SE, Q_N + 4.47\,SE]$

We can do much better if we know something about the distribution of Q_N!

Central Limit Theorem

Central Limit Theorem

Lindeberg-Levy CLT: If $\mathrm{Var}[X_i] = \sigma^2 < \infty$, then

$$\sqrt{N}(Q_N - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma)$$

Central Limit Theorem

Lindeberg-Levy CLT: If $\mathrm{Var}[X_i] = \sigma^2 < \infty$, then

$$\sqrt{N}(Q_N - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma)$$

After many samples Q_N starts to look distributed like $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{N}})$

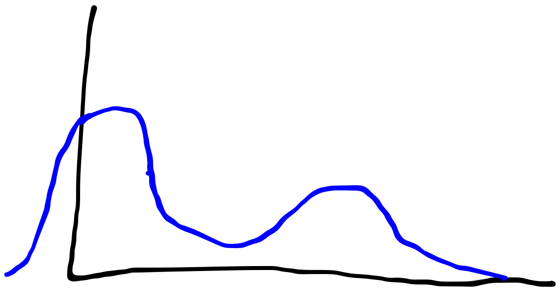
Central Limit Theorem

Lindeberg-Levy CLT: If $\mathrm{Var}[X_i] = \sigma^2 < \infty$, then

$$\sqrt{N}(Q_N - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma)$$

After many samples Q_N starts to look distributed like $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{N}})$

$$Q_1 \stackrel{D}{=} X_i$$

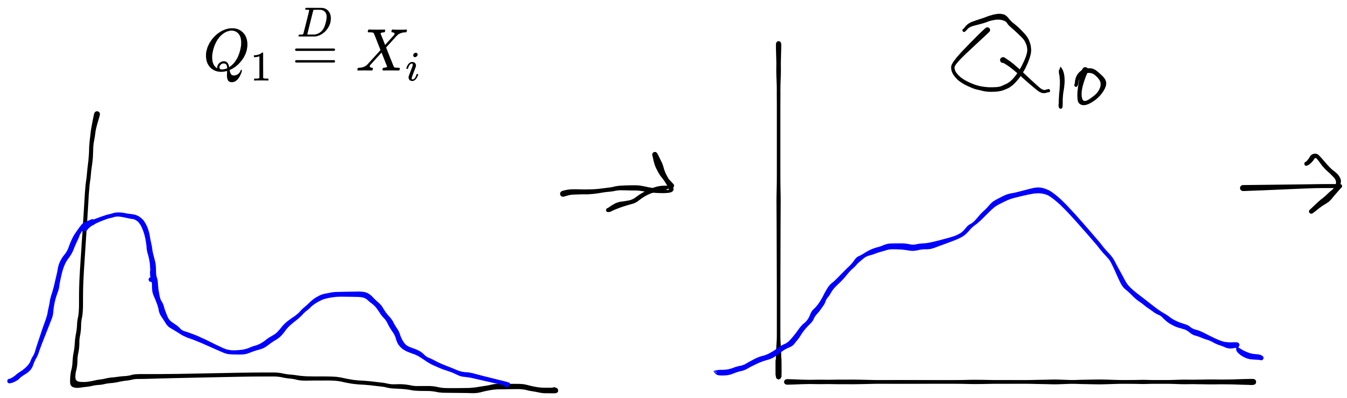


Central Limit Theorem

Lindeberg-Levy CLT: If $\mathrm{Var}[X_i] = \sigma^2 < \infty$, then

$$\sqrt{N}(Q_N - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma)$$

After many samples Q_N starts to look distributed like $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{N}})$

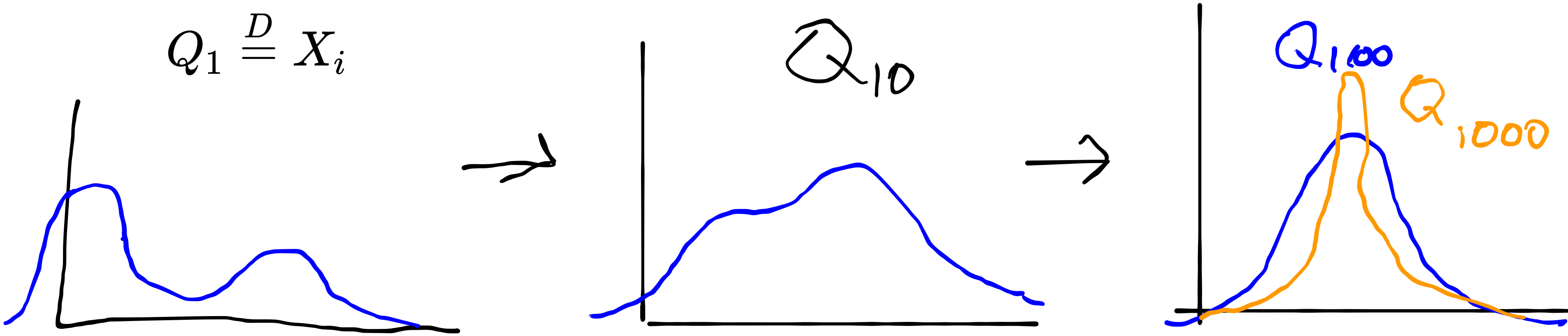


Central Limit Theorem

Lindeberg-Levy CLT: If $\mathrm{Var}[X_i] = \sigma^2 < \infty$, then

$$\sqrt{N}(Q_N - \mu) \xrightarrow{D} \mathcal{N}(0, \sigma)$$

After many samples Q_N starts to look distributed like $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{N}})$



Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

For a normal distribution,

$$P(|X - \mu| \geq t) = 1 + \mathrm{erf}\left(\frac{t - \mu}{\sqrt{2}\sigma}\right)$$

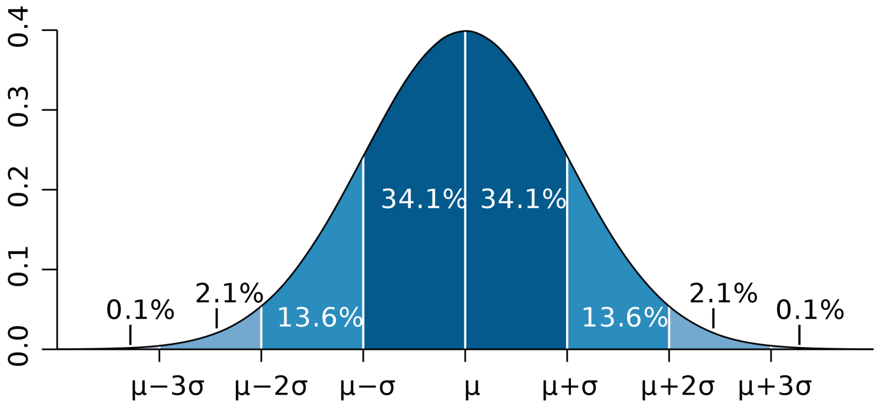
Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

For a normal distribution,

$$P(|X - \mu| \geq t) = 1 + \mathrm{erf}\left(\frac{t - \mu}{\sqrt{2}\sigma}\right)$$



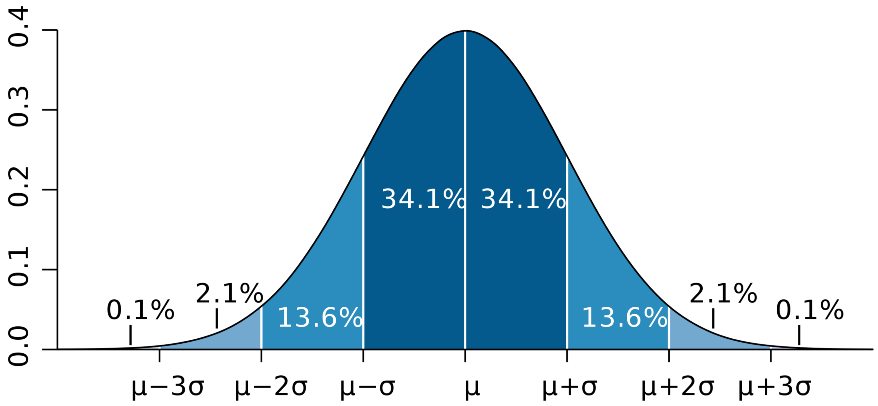
Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

For a normal distribution,

$$P(|X - \mu| \geq t) = 1 + \mathrm{erf}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)$$



$$t \approx 1.96 SE$$

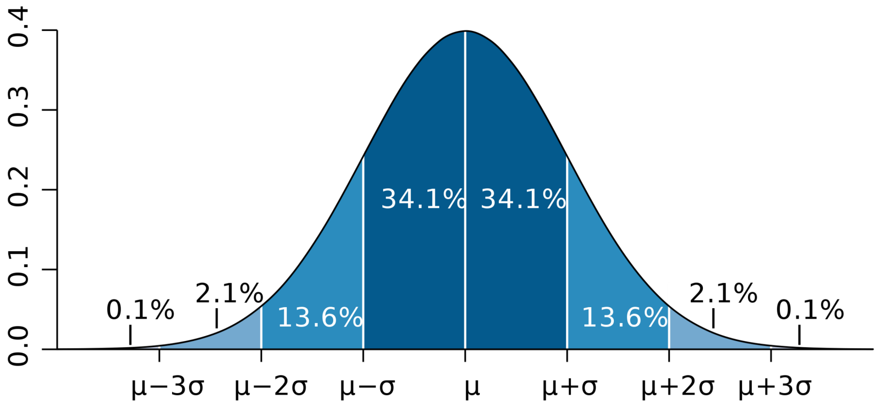
Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

For a normal distribution,

$$P(|X-\mu| \geq t) = 1 + \mathrm{erf}\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)$$



$$t \approx 1.96 SE$$

Approximate 95% CI: $[Q_N - 1.96\,SE, Q_N + 1.96\,SE]$

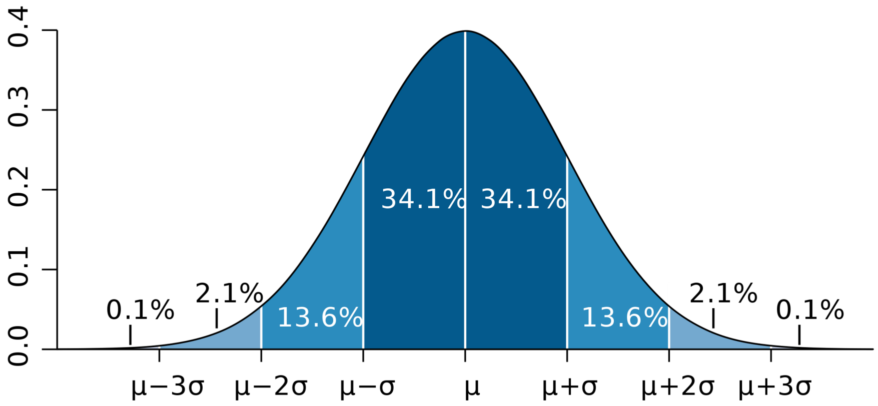
Confidence Intervals

Idea for approximate confidence interval: estimate $\mathrm{Var}(Q_N)$ with $SE^2 = \frac{s^2}{N}$ and use ~~Chebyshev~~ the central limit theorem.

Use $\gamma = 0.95$

For a normal distribution,

$$P(|X - \mu| \geq t) = 1 + \mathrm{erf}\left(\frac{t - \mu}{\sqrt{2}\sigma}\right)$$



$$t \approx 1.96 SE$$

Approximate 95% CI: $[Q_N - 1.96\,SE, Q_N + 1.96\,SE]$

(Chebyshev gave 4.47)

Importance Sampling

Want to estimate $X \sim p$ with samples from $Y_i \sim q$.

Importance Sampling

Want to estimate $X \sim p$ with samples from $Y_i \sim q$.

$$E[X] = \int x\, p(x)\, dx$$

Importance Sampling

Want to estimate $X \sim p$ with samples from $Y_i \sim q$.

$$\begin{aligned} E[X] &= \int x\, p(x)\, dx \\ &= \int x\, \frac{p(x)}{q(x)} q(x)\, dx \end{aligned}$$

Importance Sampling

Want to estimate $X \sim p$ with samples from $Y_i \sim q$.

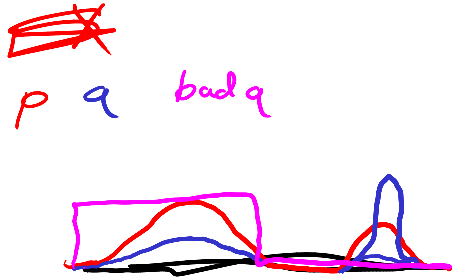
$$
\begin{aligned}
E[X] &= \int x\, p(x)\, dx \\
&= \int x \left(\frac{p(x)}{q(x)}\right) q(x)\, dx \\
&\approx \frac{1}{N} \sum Y_i \frac{p(Y_i)}{q(Y_i)}
\end{aligned}
$$

Importance Sampling

Want to estimate $X \sim p$ with samples from $Y_i \sim q$.

$$
\begin{aligned}
E[X] &= \int x\, p(x)\, dx \\
&= \int x\, \frac{p(x)}{q(x)} q(x)\, dx \\
&\approx \frac{1}{N} \sum Y_i \frac{p(Y_i)}{q(Y_i)} \\
&\approx \frac{1}{N} \sum Y_i w_i
\end{aligned}
$$

where $w_i = \frac{p(Y_i)}{q(Y_i)}$



Summary

1. Concentration Inequalities $P(X \geq t) \leq \phi(t)$

2. Law of large numbers $Q_N \rightarrow_p \mu$

3. Central Limit Theorem $Q_N \xrightarrow{D} \mathcal{N}(\mu, \frac{\sigma}{\sqrt{N}})$

4. Importance Sampling $E[X] \approx \dfrac{1}{N} \sum Y_i w_i$ where $w_i = \frac{p(Y_i)}{q(Y_i)}$