Inequalities on (expected values) (?)

1) Cauchy - Schwartz Inequality If X and Y have finite variances then

$E[XY] \le \sqrt{E(X^2)E(Y^2)}$

3) Jensen's inequality if g is convex then , convex $g(ax+(1-a)y) \le ag(x)+(1-a)g(y)$

$Eg(X) \ge g(EX)$ if concave then

$Eg(X) \le g(EX)$

3*) Hoeffdings Lemma Let $(\Omega, F, \mathbb{P})$ a probability triple and suppose that X is a $\mathbb{R}$-valued RV such that $\mathbb{P}(X \in [a,b]) = 1$, for $a<b$. Then for all $\lambda \in \mathbb{R}$

$\mathbb{E}[e^{\lambda(X-E(X))}] \le \exp\left(\frac{\lambda^2(b-a)^2}{8}\right)$

4*) Estimating p in Bernoulli: Let $X_1 \dots X_n \overset{iid}{\sim}$ Bernoulli(p) Then for $a \in (0,1)$ we have for $\delta = \frac{1}{\sqrt{n}} \cdot \sqrt{\frac{1}{2}\ln\left(\frac{2}{a}\right)}$ $\quad P(\bar{X}_n - \delta \le p \le \bar{X}_n + \delta) \ge 1-a$

* Computing multiple intervals for RVs which are not necessarily independent is quite easy using the union bound. Suppose we have m sequences of RVs $Z_1 = (X_{11}, X_{12}, \dots X_{1n})$, $Z_m = (X_{m1}, X_m\dots$ Where for each i the sequence Z_i is iid but Z_i and Z_j are not necessarily independent. Assume that each of them satisfies $P(|\frac{1}{n}\sum_{j=1}^{n} X_{ij} - E(X_{ij})| \ge \varepsilon) \le C_i$ for every i and for some number C_i. Then $P(|\frac{1}{n}\sum_{j=1}^{n} X_{ij} - E(X_{ij})| \ge \varepsilon$ for some $i) \le \sum_{i=1}^{m} C_i$ the complement is

$P(|\frac{1}{n}\sum_{j=1}^{n} X_{ij} - E(X_{ij})| < \varepsilon$ for all $i) \ge 1 - \sum_{i=1}^{m} C_i$

using $\underline{\underline{4^*}}$ $P\left(\frac{1}{n}\sum_{j=1}^{n} X_{ij} - \delta \le p_i \le \frac{1}{n}\sum_{j=1}^{n} X_{ij} + \delta \text{ for all } i\right) \ge 1-a$ where $\delta = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{2}\ln\left(\frac{2m}{a}\right)}$

This means that m appears to the logarithmic and the increase of δ with respect to m is fairly slow

This is equivalent to Bonferroni correction in multiple testing

Def.9) A $\mathbb{R}$ valued RV X is said to be sub - Gaussian with parameter λ if:

$\mathbb{E}[e^{s(X-E(X))}] \le e^{\frac{s^2\lambda^2}{2}}$ for all s ~~exact value~~

Def 10) A $\mathbb{R}$ valued RV X is said to be sub-exponential with parameter λ if:

$E[e^{s(X-E(X))}] \le e^{\frac{\lambda^2 s^2}{2}}$, for all $|s| \le \frac{1}{\lambda}$ (yields to a weaker bound)

* Theorem follow

9) Let (Ω, F, P) be a probability triple and let $X_1 \dots X_n \overset{iid}{\sim} F$ to be $\mathbb{R}$-valued sub-Gaussian RVs with parameter ~~a~~ σ then for any $\varepsilon > 0$ we get $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

$$P(\bar{X}_n - E(\bar{X}_n) \geq \varepsilon) \leq e^{-\frac{n\varepsilon^2}{2\sigma^2}}$$

10) For the sub-exponential case we get a weaker ~~bond~~ bound for the tails. The reason is because the bound on $E^{\theta X}$ only holds for small θ, the resulting estimate thus differentiates between small and big ε. We can see in the estimate below that for large ε the tail is exponential $(e^{-\varepsilon})$ & this is the reason is called sub-exponential.

Let (Ω, F, P) be a probability triple and let $X_1 \dots X_n \overset{iid}{\sim} F$ be $\mathbb{R}$-valued sub-exponential RVs with parameter λ then for any $\varepsilon > 0$ we get for $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$P(\bar{X}_n - E(\bar{X}_n) \geq \varepsilon) \leq e^{-\frac{\varepsilon^2 n}{2\lambda^2}} \vee e^{-\frac{\varepsilon n}{2\lambda}}$$

The Following Properties hold

1) Let X be a sub-Gaussian RV with parameter λ then aX is sub-Gaussian with parameter $|a|\lambda$

2) Let X be a sub-Exponential RV with parameter λ then aX is sub-exponential with parameter $|a|\lambda$

3) A sub-Gaussian RV X with parameter λ is sub-exponential with parameter λ

4) A bounded RV X, $\mathbb{P}(X \in [a,b]) = 1$ then X is sub-Gaussian with parameter $(b-a)/2$. Specifically, a Bernoulli RV is sub-Gaussian with parameter $1/2$

5) If X is sub-Gaussian with parameter λ then $Z = X^2$ is sub-exponential with parameter $8\lambda^2$

6) If X, Y independent and sub-Gaussian with parameters σ_1, σ_2 then $X+Y$ is sub-Gaussian - parameter $\sqrt{\sigma_1^2 + \sigma_2^2}$

Distribution	sub-exponential	sub-Gaussian
Gaussian	Yes	Yes
Bernoulli	Yes	Yes
Uniform	Yes	Yes
Bounded	Yes	Yes
Exponential	Yes	No
X^2	Yes	No
Weibull $(k \geq 1)$	Yes	No
Laplace	Yes	No
Pareto	No	No
Lognormal	No	No

Poisson distribution is sub-exponential. For small θ behave like a sub-Gaussian (due to the quadratic form of its MGF for small θ). Large θ: MGF grows faster → not sub-Gaussian. The tail behavior is more consistent with sub-exponential

Concentration Inequalities

1) Markov's Inequality: For a non-negative RV X, it provides a bound on the probability that X exceeds a certain value using only the expectation $E[X]$

Formula: $P(X \geq t) \leq \frac{E[X]}{t}$

When to use: Very basic inequality, useful when you only know the expectation of the RV.

2) Chebyshev's Inequality: Provide a bound on the probability that a RV deviates from its mean by more than a certain number of standard deviations.

Formula (F): $P(|X - E[X]| \geq t) \leq \frac{Var(X)}{t^2}$, $P(|Z| \geq k) \leq \frac{1}{k^2}$

When to use (wtu): When you know both the expectation and the variance of a RV.

3) Hoeffding's Inequality: A bound on the probability that the sum of bounded independent RVs deviates from its mean. For other $Y_1 \dots Y_n$ independent such that $E(Y_i) = 0$ and $a_i \leq Y_i \leq b_i$. Let $\varepsilon > 0$ and for $t > 0$ $P(\sum Y_i \geq \varepsilon) \leq e^{-t\varepsilon} \prod e^{t^2 (b_i - a_i)^2/8}$

F: For $X_1, X_2 \dots X_n$ independent and bounded (i.e. $X_i \in [a_i, b_i]$) + Bernoulli:

$P\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right] \geq t\right) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n} (b_i - a_i)^2}\right)$ | $P(|X - E(X)| \geq t) \leq 2\exp\left(\frac{-2t^2}{n(b-a)^2}\right)$

wtu: when dealing with sums of independent, bounded RVs. It gives exponentially decreasing bounds on the deviation from the mean

4) Chernoff Bounds: Similar to previous but often applied to sums of independent Bernoulli trials (binary RVs). Chernoff bounds give exponentially decreasing bounds on the tails of sums of independent RVs. $P(S_n \geq (1+\delta) E[S_n]) \leq \exp\left(-\frac{\delta^2 E[S_n]}{2+\delta}\right)$, $\delta = \frac{t}{E[S_n]}$

wtu: when you're dealing with the sum of independent RVs that are non-negative or binary

5) Bernstein's Inequality: It gives a concentration bound that takes both the variance and a bound on the individual RV into account.

F: For independent RVs $X_1, X_2 \dots X_n$ each bounded by $[X_i] \leq M$ and having variance σ^2:

$P\left(\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right] \geq t\right) \leq \exp\left(\frac{-t^2/2}{\sigma^2 + Mt/3}\right)$. M is the maximum absolute value

wtu: When working with independent RVs with known variance and upper bound of the RVs

Example 1: Let X non-negative RV with an expectation of $E[X] = 10$. Using Markov's inequality, find the upper bound on the probability that X exceeds 25

$P(X \geq 25) \leq \frac{10}{25} = 0.4$ ⇒ the probability that $X \geq 25$ is at most 0.4

Exercise 2 Suppose Y RV with mean $E[Y]$ 50 and $Var(Y) = 25$ Use Chebyshev's ... to find the upper bound on the probability that Y deviates from its mean by more than 10

$P(|X-50| \geq 10) \leq \frac{25}{10^2} = \frac{25}{100} = 0.25$ The prob that Y deviates from its mean by more than 10 is at most 0.25

Exercise 3 You toss a fair coin 100 times. Let X number of heads observed. Use Hoeffding's inequality to bound the probability that the number of heads deviates from its expected value by more than 10

Each coin toss is Bernoulli RV $X_i \in \{0,1\}$ with $P(X_i = 1) = 0.5$ (heads) and $P(X_i = 0) = 0.5$ (tails)

The expected number of heads is $E[X] = 100 \cdot 0.5 = 50$

we want to bound $P(|X-50| \geq 10)$ Let $a=0$, $b=1$, $n=100$, $t=10$

$P(|X-50| \geq 10) \leq 2 \cdot \exp\left(\frac{-2 \cdot 10^2}{100(1-0)^2}\right) = 2\exp(-2) = 2 \cdot 0.135 = 0.27$ The probability that the number of heads deviates from 50 by more than 10 is at most 0.27

Ex 4 Let $X_1, X_2, \ldots X_n$ - independent Bernoulli RVs where $X_i = 1$ with probability of 0.6 and $X_i = 0$ with probability 0.4 Define $S_n = X_1 + X_2 + \ldots + X_n$ as the sum of these variables. For $n = 100$ use Chernoff Bound to find the upper bound on the probability that S_n exceeds 70.

expected sum $E[S_n] = 100 \cdot 0.6 = 60$ $t = \text{deviation} = 70 - 60 = 10$, $\delta = \frac{t}{E(S_n)} = \frac{10}{60} = \frac{1}{6}$

$P(S_n \geq (1+\frac{1}{6}) \cdot 60) \leq \exp\left(-\frac{(\frac{1}{6})^2 \cdot 60}{2+\frac{1}{6}}\right) \Leftrightarrow P(S_n \geq 70) \leq \exp\left(-\frac{\frac{60}{36}}{\frac{13}{6}}\right) = \exp(-0.7692)$

The prob that the sum exceeds 70 is at most 0.464 ≈ 0.464

Ex 5 Consider a sequence of independent RVs $X_1, X_2 \ldots X_5$ where each X is bounded such that $X_i \in [-1, 1]$ and has variance $\sigma_i^2 = 0.25$ Use Bernstein's inequality to bound the probability that the sum $S = \sum_{i=1}^{5} X_i$ deviates from the mean by at least 2

$P(S - E(S) \geq t) \leq \exp\left(\frac{-t^2/2}{\sigma^2 + Mt/3}\right)$ M in this case is 1

$\sigma^2 = \sum_{i=1}^{5} 5\sigma_i^2 = 5 \cdot 0.25 = 1.25$

$e^{-1.04}$

$P(S - E(S) \geq 2) \leq \exp\left(\frac{-2^2/2}{1.25+2/3}\right) = \exp\left(\frac{-2}{1.92}\right) = \exp(-1.04) \approx 0.353$

The prob that the sum deviates from its mean by at least 2 is 0.353

6) Mill's Inequality: Let $Z \sim N(0,1)$. Then $\mathbb{P}(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}$

works for Normal Random Variables

Distribution

Distribution	Mean	Variance
Point Mass at a	a	0
Bernoulli(p)	p	$p(1-p)$
Binomial (n, p)	np	$np(1-p)$
Geometric (p)	$1/p$	$(1-p)/p^2$
Poisson (λ)	λ	λ
Uniform (a, b)	$(a+b)/2$	$(b-a)^2/12$
Normal (μ, σ^2)	μ	σ^2
Exponential (β)	β	β^2
Gamma (a, β)	$a\beta$	$a\cdot\beta^2$
Beta (a, β)	$a/(a+\beta)$	$a\beta/((a+\beta)^2(a+\beta+1))$
t_v	0 if $v>1$	$v/(v-2)$ (if $v>2$)
χ^2_p	p	$2p$

Exercise 1: Consider 100 independent coin flips. We want to find an upper bound on the prob. that the number of heads is greater of equal than 75. (use Markov's & Chebyshev's)

Markov's $E(X) = n\cdot p = 100 \cdot \frac{1}{2} = 50 \quad P(X \geq 75) \leq \frac{E(X)}{75} = \frac{50}{75} = 0.666$

Chebyshev's: $E(X) = n\cdot p = 100\,\frac{1}{2} = 50$, $Var(X) = n\cdot p(1-p) = 100\,\frac{1}{2}\,\frac{1}{2} = 25$

$P(|X-50| \geq 25) \leq 25/(25)^2 = 0.04$ → better than Markov's

Chernoff Bound $P(S_n \geq (1+\delta)E[S_n]) \leq \exp\left(-\frac{\delta^2 E[S_n]}{2+\delta}\right)$, $\delta = \frac{t}{E[S_n]}$

$\mu = E[S_n] = \sum_{i=1}^{100} E[X_i] = 100\,\frac{1}{2} = 50$

$S \geq 75$ (deviation of 75 heads from the expected 50) $\delta = \frac{75-50}{50} = \frac{25}{50} = \frac{1}{2}$

$P(S_n \geq \frac{3}{2}E[S_n]) \leq \exp\left(\frac{-\frac{1}{4}\,50}{2+\frac{1}{2}}\right) = \exp\left(\frac{-\frac{25}{2}}{\frac{5}{2}}\right) = \exp(-5) \simeq 0.0067$

which of the following will exponentially concentrate i.e. for some C_1, C_2, C_3, C_4

$P(Z - E[Z] \geq \varepsilon) \leq \underbrace{C_1 e^{-C_2 n\varepsilon^2}}_{\text{for small deviation}} \vee \underbrace{C_3 e^{-C_4 n(\varepsilon+1)}}_{\text{for large deviations}}$

1) The empirical variance of iid RVs with finite mean?

Let $X_1, X_2, \ldots X_n$ be iid with mean $\mu = E[X_i]$ and $\sigma^2 = E[(X_i-\mu)^2]$. The empirical variance of these RVs is given: $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, sample mean.

true variance σ^2 to show concentration we want to bound the probability that the empirical variance deviates significantly from σ^2: $P(S_n - \sigma^2 \geq \varepsilon)$, $\varepsilon > 0$

$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = M_n - \bar{X}^2$, $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ second moment of the sample

→ We usually use Chernoff's inequality when we are dealing with sums of independent random variables. We need to bound the probability of large deviations of these sums from their mean

→ A RV X is sub-Gaussian if its tails decay at least as fast as the tails of a Gaussian distrib.

X is sub-Gaussian if there exists a constant $k > 0$ such that for all $t > 0$

$P(|X - E[X]| \geq t) \leq 2\exp\left(-\frac{t^2}{2k^2}\right)$

...ntration Inequalities.

(1) Markov's Inequality: Let $X \in \mathcal{L}^1(P)$ non-negative $\mathbb{R}$-valued RV

$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$ for any $\varepsilon > 0$

Proof: let $A_\varepsilon = [\varepsilon, +\infty)$: $I_{A_\varepsilon}(x) + I_{A_\varepsilon^c}(x) = 1$ — indicator function of set A_ε

$A_\varepsilon = \{X \geq \varepsilon\}$ ↳ set of outcomes $\omega \in \Omega$ which the RV $X(\omega)$ is at least ε

~~Note~~ $I_{A_\varepsilon}(X)$: indicator function that equals 1 when $X \geq \varepsilon$ and 0 otherwise

$I_{A_\varepsilon^c}(X)$ —— // —— $X < \varepsilon$ —— // ——

For set A_ε we know: $X \geq \varepsilon \Rightarrow X I_{A_\varepsilon}(X) \geq \varepsilon I_{A_\varepsilon}(X)$ X is at least ε if replace x with ε

$E[X] \geq E[\varepsilon I_{A_\varepsilon}(X)]$ since ε constant we can factor it out of the expectation:

$E[X] \geq \varepsilon E[I_{A_\varepsilon}(X)]$ The expectation of an indicator function is the probability of the set A_ε: $E(X) \geq \varepsilon \cdot P(X \geq \varepsilon) \Leftrightarrow$ ~~P~~ $P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}$

(2) Chebyshev's Inequality: For any RV X and any $\varepsilon > 0$

$P(|X| \geq \varepsilon) \leq \frac{E[|X|]}{\varepsilon}$

using Markov's Inequality: $P(|X| \geq \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$ apply this inequality for X^2:

$P(X^2 \geq \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}$, $P(|X| \geq \varepsilon) = P(X^2 \geq \varepsilon^2)$ → uses the second moment of X (this gives more info about how spread out X is)

For Chebyshev's Inequality we look at the deviation from the mean:

$P(|X - E(X)| \geq \varepsilon) = P((X - E(X))^2 \geq \varepsilon^2)$ using Markov's Inequality:

$P((X - E(X))^2 \geq \varepsilon^2) \leq \frac{E((X - E(X))^2)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$

I.I.D: 1) Independent: when RVs and the outcome of the one doesn't affect the outcome of another: $P(X_1 \in A \text{ and } X_2 \in B) = P(X_1 \in A) P(X_2 \in B)$

2) Identically distributed: RV follow the same probability distribution $P(X_1 \leq x) = P(X_2 \leq x)$ (same shape & parameters)