Probability and Random Process

COSE382

Wonzoo Chung

Inequalities

Theorem 10.1.1 (Cauchy-Schwarz inequality). For any r.v.s X and Y with finite variances,

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}.$$

Proof. For any t, we have

$$0 \leq E(Y - tX)^2 = E(Y^2) - 2tE(XY) + t^2 E(X^2).$$

This quadratic is minimized at $t = E(XY)/E(X^2)$. Furthermore, at this t

$$\begin{aligned} 0 &\leq E(Y^2) - 2E(XY)^2/E(X^2) + E(XY)^2/E(X^2) \\ 0 &\leq E(Y^2) - E(XY)^2/E(X^2) \\ E(XY)^2 &\leq E(X^2)E(Y^2) \end{aligned}$$

Note that when X and Y are zero-mean Cauchy-Schwarz inequality implies

$$\left(\frac{E(XY)}{\sqrt{E(X^2)}\sqrt{E(Y^2)}}\right)^2 = \left(\frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}\right)^2 = \rho_{XY}^2 \leq 1$$

Example 10.1.3 Let X be a nonnegative r.v. ($X \geq 0$). We want an upper bound on $P(X = 0)$.

Let $I_{X>0}$ be the indicator r.v. of $X > 0$. Then $X = X \cdot I_{X>0}$ and

$$E(X) = E(X \cdot I_{X>0}) \leq \sqrt{E(X^2)E(I_{X>0})}.$$

Since $E(I_{X>0}) = P(X > 0)$, we have $P(X > 0) \geq \dfrac{(EX)^2}{E(X^2)}$, and

$$P(X = 0) = 1 - P(X > 0) \leq 1 - \frac{(EX)^2}{E(X^2)} = \frac{\mathrm{Var}(X)}{E(X^2)}.$$

For example, let $X = I_1 + \cdots + I_n$, where the I_j are uncorrelated indicator r.v.s. with $p_j = E(I_j)$.

$$\mathrm{Var}(X) = \sum_{j=1}^{n} \mathrm{Var}(I_j) = \sum_{j=1}^{n}(p_j - p_j^2) = \sum_{j=1}^{n} p_j - \sum_{j=1}^{n} p_j^2 = \mu - c,$$

where $\mu = E(X)$, $c = \sum_{j=1}^{n} p_j^2$. Also, $E(X^2) = \mathrm{Var}(X) + (EX)^2 = \mu^2 + \mu - c$. So

$$P(X = 0) \leq \frac{\mathrm{Var}(X)}{E(X^2)} = \frac{\mu - c}{\mu^2 + \mu - c} \leq \frac{1}{\mu + 1},$$

Bounds on tails

Theorem 10.1.10 (Markov inequality). For any r.v. X and constant $a > 0$,

$$P(|X| \geq a) \leq \frac{E|X|}{a}.$$

Proof. Let $Y = \dfrac{|X|}{a}$. We need to show that $P(Y \geq 1) \leq E(Y)$. Let $I_{Y \geq 1}$ denote the indicator r.v.

$$I_{Y \geq 1}(s) = \begin{cases} 1 & Y(s) \geq 1 \\ 0 & \text{else} \end{cases}$$

Then we have

$$I_{Y \geq 1} \leq Y.$$

Taking the expectation of both sides, we have Markov's inequality.

$$E(I_{Y \geq 1}) \leq E(Y) \implies P(Y \geq 1) \leq E(Y)$$

Theorem 10.1.11 (Chebyshev). Let X have mean μ, and variance σ^2. Then for any $a > 0$,

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Proof. By Markov's inequality,

$$P(|X - \mu| \geq a) = P((X - \mu)^2 \geq a^2) \leq \frac{E(X - \mu)^2}{a^2} = \frac{\sigma^2}{a^2}.$$

Theorem 10.1.12 (Chernoff). For any r.v. X and constants $a > 0$ and $t > 0$,

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

Proof. The function $g(x) = e^{tx}$ is strictly increasing. So by Markov's inequality,

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

Exmaple 10.1.13 (Bounds on a Normal tail). $P(|Z| > 3) = 2\Phi(-3) \approx 0.003$. Let's see what upper bounds are obtained from Markov's, Chebyshev's, and Chernoff's inequalities.

- Markov: $E|Z| = \sqrt{2}/\pi$. Then

$$P(|Z| > 3) \leq \frac{E|Z|}{3} = \frac{1}{3} \cdot \sqrt{\frac{2}{\pi}} \approx 0.27.$$

- Chebyshev:

$$P(|Z| > 3) \leq \frac{1}{9} \approx 0.11.$$

- Chernoff (after using symmetry of the Normal):

$$P(|Z| > 3) = 2P(Z > 3) \leq 2e^{-3t}E(e^{tZ}) = 2e^{-3t} \cdot e^{t^2/2},$$

using the MGF of the standard Normal distribution. The right-hand side is minimized at $t = 3$, as found by setting the log-derivative equal to 0,

$$P(|Z| > 3) \leq 2e^{-9/2} \approx 0.022.$$

Law of large numbers

Let $X_1, X_2, X_3, \cdots$ be i.i.d. with mean μ and variance σ^2. Then the sample mean

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

is a r.v. with $E(\bar{X}_n) = \mu$ and $\mathrm{Var}(X_n) = \dfrac{\sigma^2}{n}$. Then we have

Theorem 10.2.1 (Strong law of large numbers). The sample mean $\bar{X}_n$ converges to the true mean μ pointwise as $n \to \infty$, with probability 1.

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

Theorem 10.2.2 (Weak law of large numbers). For all $\epsilon > 0$, .

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Proof. Fix $\epsilon > 0$. By Chebyshev's inequality,

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

As $n \to \infty$, the right-hand side goes to 0, and so must the left-hand side.

Example 10.2.3 (Running proportion of Heads). Consider tossing a fair coin independently many times. We are interested in the proportion of Heads as the number of toss increases.

Let X_i be the indicator of Head in the ith toss. Then $X_1, X_2, \cdots$ be i.i.d. Bern(1/2) and $\bar{X}_n$ denotes the sample mean, the proportion of Heads, up to n tosses.

> **SLLN:** As n increases $\bar{X}_n$ eventually converges to $1/2$, with probability 1.
>
> **WLLN:** For any $\epsilon > 0$, the probability of $\bar{X}_n$ being more than ϵ away from $1/2$ can be made as small as we like by letting n grow.

As $n \to \infty$, $P(\text{number of } H = n/2) \to 1$?

$$
\begin{aligned}
P(\text{number of } H = n/2) &= \frac{\binom{n}{\frac{n}{2}} \frac{1}{2^n}}{2^n \frac{1}{2^n}} = \frac{\binom{n}{\frac{n}{2}}}{2^n} = \frac{n!}{\frac{n}{2}!\frac{n}{2}!2^n} \\
\log P(\text{number of } H = n/2) &= \log n! - 2\log\frac{n}{2}! - n\log 2 \\
&\approx n\log n - n - 2\left(\frac{n}{2}\log\frac{n}{2} - \frac{n}{2}\right) - n\log 2 \\
&= n\log n - n\log n + n\log 2 - n\log 2 = 0
\end{aligned}
$$

We have used Stirling's approximation $n! \to n\log n - n$ as $n \to \infty$.

Central Limit Theorem

Theorem 10.3.1 (Central limit theorem). Let $X_1, X_2, \cdots$, be i.i.d. with mean μ and variance σ^2, and $\bar{X}_n = \frac{1}{n}(X_1 + \cdots X_n)$. As $n \to \infty$,

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \to \mathcal{N}(0,1) \text{ in distribution.}$$

Proof. Without loss of generality, $\mu = 0$, $\sigma^2 = 1$. Let $M(t) = E(e^{tX_j})$. Then $M(0) = 1, M'(0) = \mu = 0$, and $M''(0) = \sigma^2 = 1$.

$$E(e^{t(X_1+\cdots+X_n)/\sqrt{n}}) \quad = \quad \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n, \quad \left(\to e^{t^2/2} \text{ we will show }\right)$$

$$\begin{aligned}
\lim_{n\to\infty} n \log M\left(\frac{t}{\sqrt{n}}\right) &= \lim_{y\to 0} \frac{\log M(yt)}{y^2} && \text{where } y = 1/\sqrt{n} \\
&= \lim_{y\to 0} \frac{tM'(yt)}{2yM(yt)} && \text{by L'Hôpital's rule} \\
&= \frac{t}{2}\lim_{y\to 0}\frac{M'(yt)}{y} = \frac{t^2}{2}\lim_{y\to 0} M''(yt) = \frac{t^2}{2}.
\end{aligned}$$

Hence, $(M(t/\sqrt{n}))^n$, the MGF of $\sqrt{n}\bar{X}_n$, approaches $e^{t^2/2}$, the MGF of $\mathcal{N}(0,1)$.

Example 10.3.4 (Poisson convergence to Normal). Let $Y \sim \text{Pois}(n)$. We can consider Y to be a sum of n i.i.d. Pois(1) r.v.s. Therefore, for large n,

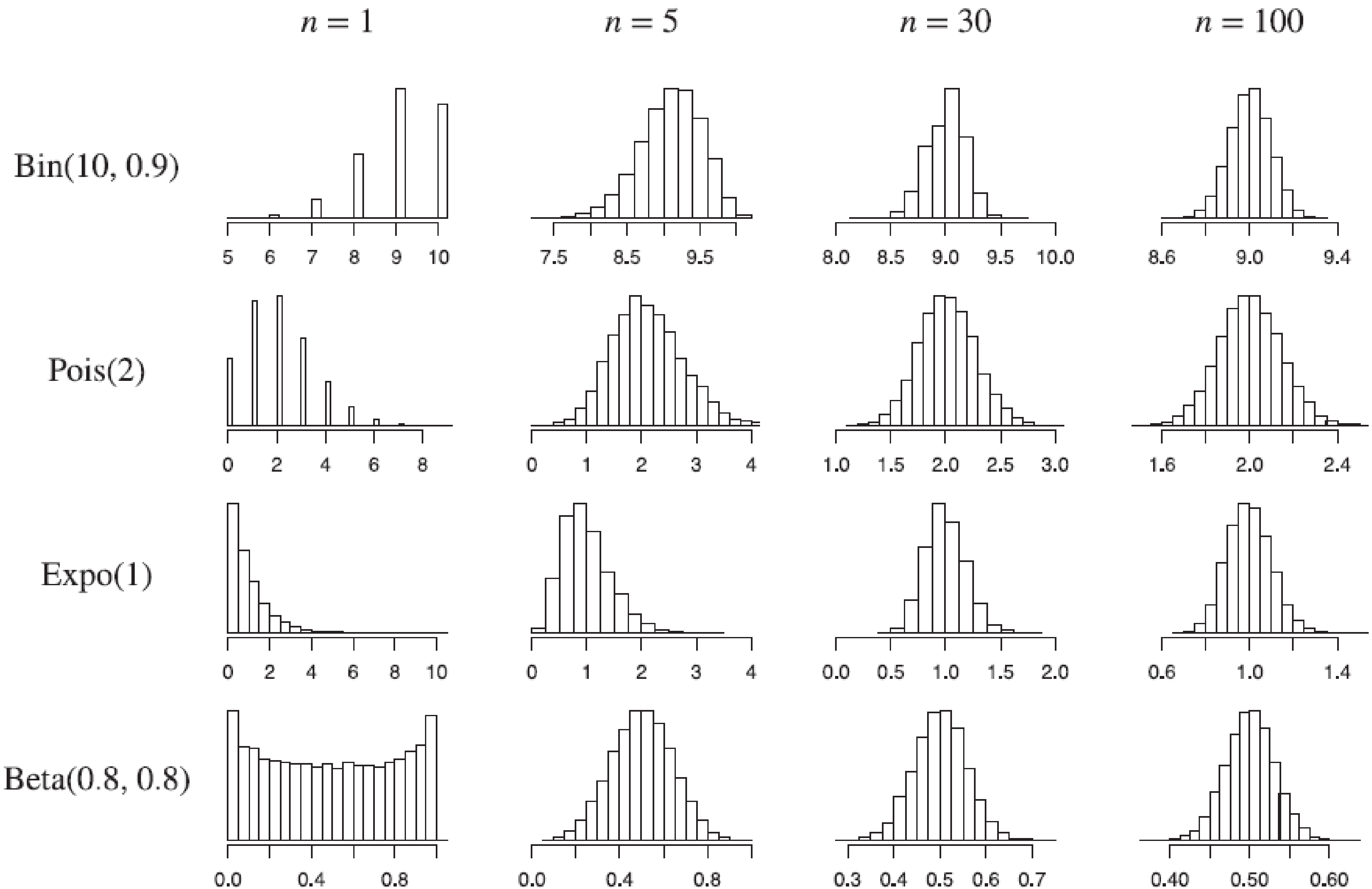
$$Y \sim \mathcal{N}(n, n).$$

Example 10.3.5 (Gamma convergence to Normal). Let $Y \sim \text{Gamma}(n, \lambda)$. We can consider Y to be a sum of n i.i.d. Expo(λ) r.v.s. Therefore, for large n,

$$Y \sim \mathcal{N}(\frac{n}{\lambda}, \frac{n}{\lambda^2}).$$

Theorem 10.3.6 (Binomial convergence to Normal). Let $Y \sim \text{Bin}(n, p)$. We can consider Y to be a sum of n i.i.d. Bern(p) r.v.s. Therefore, for large n,

$$Y \sim \mathcal{N}(np, np(1-p)).$$



$n = 1$
$n = 5$
$n = 30$
$n = 100$
Bin(10, 0.9)
Pois(2)
Expo(1)
Beta(0.8, 0.8)
5 6 7 8 9 10
7.5 8.5 9.5
8.0 8.5 9.0 9.5 10.0
8.6 9.0 9.4
0 2 4 6 8
0 1 2 3 4
1.0 1.5 2.0 2.5 3.0
1.6 2.0 2.4
0 2 4 6 8 10
0 1 2 3 4
0.0 0.5 1.0 1.5 2.0
0.6 1.0 1.4
0.0 0.4 0.8
0.0 0.4 0.8
0.3 0.4 0.5 0.6 0.7
0.40 0.50 0.60