LECTURE 13

BOUNDING INEQUALITIES

→ $P(X \geq a)$ is called the tail probability.

In practice, sometimes we do not know the distribution X, but might know its mean and variance.

→ Today, we will provide upper bounds for tail probabilities that hold for any distribution. A few of these are:

→ Markov Inequality → Chebyshev Inequality

→ Chernoff Bound.

→ If time permits, we will also cover generating RVs | distributions

MARKOV INEQUALITY

→ Suppose X is a non-negative RV (can be discrete or continuous).

Then, $\boxed{P(X \geq a) \leq \frac{E[X]}{a}}$ for $a > 0$.

Note that we are only using the knowledge of $E[X]$.

Why is this true?

Proof: Consider X a continuous RV which is non-negative by assumption.

Then $E[X] = \int_0^\infty x f_X(x)dx = \underbrace{\int_0^a x f_X(x)dx}_{\text{non-negative}} + \int_a^\infty x f_X(x)dx$

$\geq \int_a^\infty x f_X(x)dx$ (dropping 1st term)

$\geq \int_a^\infty a f_X(x)dx$ (replacing x by a)

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

$\Rightarrow P(X \geq a) \leq \frac{E[X]}{a}$ (rearranging terms).

CHEBYSHEV INEQUALITY

→ Suppose now we know both the 1st and 2nd moments of a given unknown distribution. This is equivalent to saying that we know its mean and variance as $Var(X) = \underbrace{E[X^2]}_{\text{2nd Moment}} - \underbrace{(E[X])^2}_{\text{1st Moment}}$

→ Let m be the mean and let σ^2 be the variance of X.

Then, $\boxed{P(|X-m| \geq a) \leq \frac{\sigma^2}{a^2}}$ No restrictions on a

→ Note that now, there is no restriction on X as well.

Proof: Let $g(X) = (X-m)^2$ and $Y = g(X)$.

Then, Y is non-negative so we can apply the Markov inequality to Y.

Then, $P(Y \geq b) \leq \frac{E[Y]}{b}$, $b > 0$

$$\Rightarrow P(g(X) \geq a^2) \leq \frac{E[g(X)]}{a^2}, \quad a^2 > 0 \qquad (b = a^2)$$

$$\Rightarrow P((X-m)^2 \geq a^2) \leq \frac{E[(X-m)^2]}{a^2}$$

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

We just proved the Chebyshev Inequality as a corollary to the Markov Inequality.

Example $N(0,1)$.

Bound $P(|X| \geq 2)$ using Chebyshev inequality) where $X \sim N(0,1)$.

WKT, $P(|X-m| \geq a) \leq \frac{\sigma^2}{a^2}$ from the Chebyshev Inequality

$m = 0$ and $\sigma^2 = 1$ as $X \sim \mathcal{N}(0,1)$.

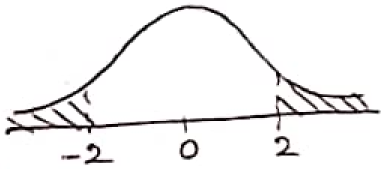
$\therefore$, by substitution,

$$P(|X| \geq 2) \leq \frac{1}{4}.$$

Let's compare to Q-function calculations:

$$P(|X| \geq 2) = P(X \geq 2) + P(X \leq -2)$$

$$= 2\,Q(2) \quad (\text{by symmetry})$$



$$= 2 \times 0.0226$$

$$\simeq 0.045 << \frac{1}{4}.$$

Clearly, the Chebyshev inequality, although holds, is not exactly a good bound in this case.

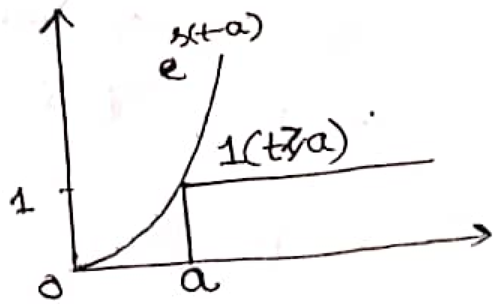
CHERNOFF BOUND

→ Consider $a > 0$.

$$P(X \geq a) = \int_{-\infty}^{\infty} 1(t \geq a)\, f_X(t)\, dt \quad \text{where } 1(t \geq a) = \begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

When $t = a$, $e^{s(t-a)} = 1$.

When $t = 0$, $e^{s(t-a)} = e^{-sa}$.



$$P(X \geq a) = \int_{-\infty}^{\infty} 1(t \geq a)\, f_X(t)\, dt$$

$$\leq \int_{-\infty}^{\infty} e^{+s(t-a)} f_X(t)\, dt$$

$$= e^{-sa} \int_{-\infty}^{\infty} e^{st} f_X(t)\, dt$$

$$= e^{-sa} E[e^{sX}]$$ → this is true for any positive s.

Does $E[e^{sX}]$ remind you of something we have seen already?

Yes! It is very similar to $\Phi_X(\omega) = E[e^{j\omega X}]$.

<u>Example</u> $X \sim \mathcal{N}(0,1)$.

$$\mathbb{E}[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x^2-2sx+s^2)}{2} + \frac{s^2}{2}} \, dx \quad \text{(computing the square)}$$

$$= e^{s^2/2} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-s)^2}{2}} \, dx}_{\int_{-\infty}^{\infty} \mathcal{N}(s,1) = 1}$$

$$= e^{s^2/2}.$$

$\therefore$, $P(X \geq a) \leq e^{-sa} e^{s^2/2}$

This is true for any $s > 0$.

Consider $s_1 = P(X \geq a) \leq e^{-s_1 a} e^{s_1^2/2}$
$s_2 = P(X \geq a) \leq e^{-s_2 a} e^{s_2^2/2}$ 2 different upper bounds.

Thus, this should also be true for the best s.

We have $P(X \geq a) \leq e^{h(s)}$ where $h(s) = \frac{s^2}{2} - as$.

To find the best s, we differentiate the RHS and set it to 0 and solve for 's'. But as the exponential function is monotonic, minimizing $e^{h(s)}$ = minimizing $h(s)$.

$$\therefore, \frac{d}{ds}(h(s)) = 0 \Rightarrow \frac{2s}{2} - a = 0$$

$$\Rightarrow s = a.$$

$\therefore$, $P(X \geq a) \leq e^{-a^2/2}$ for the Gaussian $X \sim \mathcal{N}(0,1)$.

Clearly, the bigger the a, the lesser the probability. So this is a very useful upperbound.

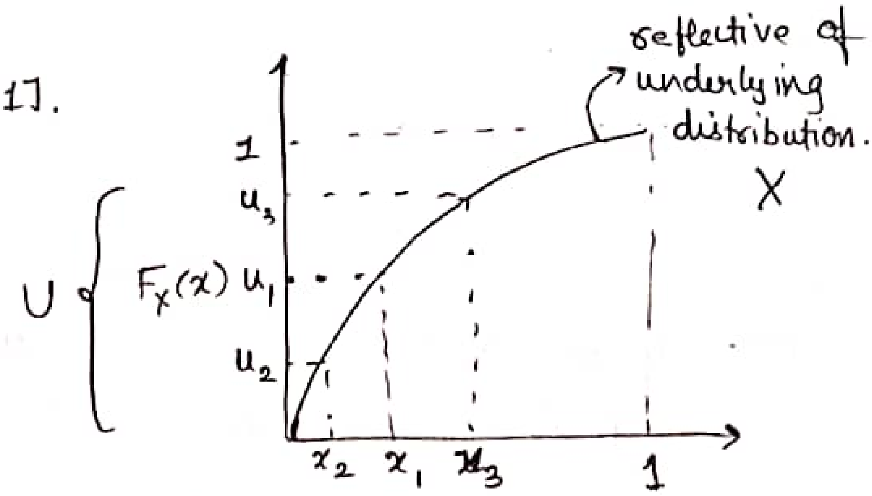
GENERATING RANDOM VARIABLES

→ Suppose we wish to generate a RV X that is continuous and with CDF $F_X(x)$.

Consider $U \sim$ Uniform $[0, 1]$.

Compute $P(X \le x)$.

We generate values uniformly at random from 0 to 1.



We have $x_1 = F_X^{-1}(u_1)$ where $u_1 \sim$ Uniform $[0, 1]$

$$\therefore\ P(X \le x) = P(F_X^{-1}(U) \le x)$$

$$= P(U \le F_X(x))$$

$$= F_X(x)$$

as $P(X \le a) = a$ for $X \sim U[0, 1]$, $0 \le a \le 1$.

Procedure

1) Generate $U \sim$ Uniform $[0, 1]$.
2) Let $X = F_X^{-1}(U)$.

Example

→ Let us generate an exponential RV with parameter λ.

Recall : $X \sim \exp(\lambda) \longrightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & \text{otherwise} \end{cases}$

and $P(X \ge x) = e^{-\lambda x}$

(a) Generate $U \sim$ Uniform $[0, 1]$.

2) Compute F_X^{-1}.

$$F_X(x) = 1 - e^{-\lambda x} = u$$

$$\Rightarrow 1 - u = e^{-\lambda x}.$$

$$\Rightarrow \ln(1-u) = -\lambda x$$

$$\Rightarrow x = -\frac{1}{\lambda}\ln(1-u)$$

Also, if $U \sim$ uniform $[0, 1]$, then what can you say about $1-U$?

- $1-U$ is also uniform in $[0, 1]$.

$\therefore$, $x = -\frac{1}{\lambda}\ln(u)$.

To generate samples,

$$x_1 = -\frac{1}{\lambda}\ln(u_1) \qquad x_2 = -\frac{1}{\lambda}\ln(u_2) \quad \text{and so on.}$$

What happened here with the negative sign?

$\ln()$ is negative as $0 \leq u \leq 1$. $\therefore$, the $-$ cancel out and finally yield a positive number.

BOX MULLER METHOD

Example But what if the distribution has a CDF which is hard to invert? Suppose $X \sim N(0,1)$ and we want to generate samples from X that uses $U[0, 1]$.

Recall from last class: $X_1 = R\cos\theta$

$X_2 = R\sin\theta$

We had $\left[\int f_X(x)dx\right]\left[\int f_Y(y)dy\right] = 1$

and changed variables X and Y to R and θ.

We also noted that $\theta \sim$ Uniform $[0, 2\pi]$

$\frac{R^2}{2} \sim \exp(1)$.

Then,

$$X_1 = \sqrt{-2\ln(U1)}\ \cos(2\pi U2).$$

$$X_2 = \sqrt{-2\ln(U2)}\ \cos(2\pi U1).$$

where U_1 and U_2 are independent and uniformly distributed in $[0, 1]$.

We generate 2 Gaussians $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$.