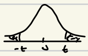
Week 11

**Thm.** (tighter inequality for normal tails)

If $Z \sim N(0,1)$, then for $t>0$: $\boxed{P(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}}$




<u>Remark</u>: If $t=2$, $P(|Z| \geq 2) \doteq \sqrt{2/\pi}\, \frac{e^{-2}}{2} = 0.054 << 1/4$ (from above chebyshev)

<u>Corollary</u>: In general, $\underbrace{\sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}}_{\text{exponentially small}} << \underbrace{\frac{1}{t^2}}_{\text{polynomially true}}$

<u>Proof.</u> $P(|Z| \geq t) = 2P(Z \geq t)$ by symmetry of $f_Z$

$$= 2\int_t^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, dz$$

$$\leq 2\int_t^{\infty} \frac{1}{\sqrt{2\pi}} \frac{z}{t} e^{-\frac{1}{2}z^2}\, dz$$

$$= \frac{2}{\sqrt{2\pi}} \frac{1}{t} e^{-\frac{1}{2}z^2} \Big|_{\infty}^{t}$$

$$= \sqrt{\frac{2}{\pi}} \frac{1}{t} e^{-\frac{1}{2}t^2}. \quad \square$$

<u>Poisson (identity)</u>: If $X \sim \text{Poisson}(\lambda)$, then $f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$

So, $\boxed{\begin{cases} P(X=0) = e^{-\lambda} \\ P(X=x+1) = e^{-\lambda} \frac{\lambda^{x+1}}{(x+1)!} \\ \qquad = P(X=x) \frac{\lambda}{x+1} \end{cases}}$

<u>note</u>: Recursion-like identities (like this one) may be useful in various situations that require such computations

<u>Gamma (identity)</u>: If $X_{\alpha,\beta} \sim \text{Gamma}(\alpha,\beta)$ with $\begin{vmatrix} f(x|\alpha,\beta) \\ \alpha>1, \beta>0 \end{vmatrix}$

Then $\forall\, a,b$

$\boxed{P(X_{\alpha,\beta} \in (a,b)) = \beta(f(a|\alpha,\beta) - f(b|\alpha,\beta)) + P(X_{\alpha-1,\beta} \in (a,b))}$

<u>Proof.</u> (Application of IBP)

<u>Remark</u>: If $\alpha \in \mathbb{N}$, the above identity allows us to recurse to $\text{Gamma}(\alpha-1,\beta)$; $\text{Gamma}(\alpha-2,\beta), \ldots$, and so on to $\text{Gamma}(1,\beta) = \text{Exp}(\beta)$.

<u>Stein's Identity for Normal RV</u>: If $X \sim N(\theta, \sigma^2)$,

$g$ is a differentiable function s.t. $E|g'(x)| < \infty$

Then

$\boxed{E[g(X)(X-\theta)] = \sigma^2 E(g'(X))}$

<u>Proof.</u> $\text{LHS} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} g(x)(x-\theta) e^{\frac{-1}{2\sigma^2}(x-\theta)^2}\, dx$

IBP?

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} g(x)(-\sigma^2)\, d\left(e^{\frac{-1}{2\sigma^2}(x-\theta)^2}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \left( \cancel{g(x)(-\sigma^2) e^{\frac{-1}{2\sigma^2}(x-\theta)^2} \Big|_{-\infty}^{\infty}} + \int_{-\infty}^{\infty} \sigma^2 e^{\frac{-1}{2\sigma^2}(x-\theta)^2}\, dg(x) \right)$$

0

review more details

$$= \sigma^2 E(g'(X)). \quad \square$$

<u>Application</u>: we know $E(X) = \theta$

$E^2(X) = [E(X)]^2 + \text{var}(X)$

$= \theta^2 + \sigma^2$

By Stein's lemma,

$E(X^3) = E(X^2(X-\theta+\theta))$

$= E(X^2(X-\theta)) + \theta E(X^2)$

how?

choosing $g(x) = x^2$

$= \sigma^2 E(2X) + \theta(\theta^2 + \sigma^2)$

$= 2\sigma^2\theta + \theta^3 + \theta\sigma^2$

$= \theta^3 + 3\sigma^2\theta$

and so on, for $E(X^n)$, $n = 4, 5, \dots$

<u>Hwang's Identity</u> (for Discrete variables)

Let $g$: function with $|Eg(X)| < \infty$ and $|g(-1)| < \infty$.

If $X \sim \text{Poisson}(\lambda)$, then $\underline{E(\lambda g(X)) = E(X g(X-1))}$

If $X \sim \text{NegBinom}(r, p)$, then $\underline{E((1-p)g(X)) = E\left[\frac{X}{r+X-1} g(X-1)\right]}$

<u>Application</u>: • If $X \sim \text{Poisson}(\lambda)$, then $E(X) = V(X) = \lambda$.

Thus, $E(\lambda X^2) = EX(X-1)^2$, letting $g(x) = x^2$

$\lambda(\lambda^2 + \lambda) = E(X^3 - 2X^2 + X)$

$\lambda^3 + \lambda^2 = EX^3 - 2(\lambda^2 + \lambda) + \lambda$

$\Rightarrow EX^3 = \lambda^3 + 3\lambda^2 + \lambda$

• If $X \sim \text{NegBinom}(r, p)$, take $g(x) = r + x$.

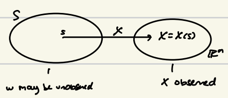
Then $E(1-p)(r+X) = E\frac{X}{r+X-1}(r+X-1)$

$\Rightarrow (1-p)r + (1-p)EX = EX$

$\Rightarrow EX = \frac{(1-p)r}{p}$.

## 4.1: Joint and marginal distributions

<u>Def</u>: An n-dimensional random vector is a function from a sample space into space $\mathbb{R}^n$ (may change to other spaces)




<u>Ex</u>: $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$

associated w/ msmt's of a (random) person

where $X_1$ = temp.

$X_2$ = height

$X_3$ = bp

$\vdots$

<u>note</u>: If $n=2$, $X = (X_1, X_2) \in \mathbb{R}^2$ is called a <u>bivariate</u> vector

<u>Ex</u>: consider bivariate RV $(X,Y) \in \mathbb{R}^2$ corresponding to the following

expt (tossing two fair dice): $\begin{cases} X = \text{sum of two dice} \\ Y = \text{absolute diff. of two dice} \end{cases}$

Write sample point $s \in S$ as $s = (s_1, s_2)$ for $s_1, s_2 \in \{1, \dots, 6\}$

then $\begin{cases} X = s_1 + s_2 \\ Y = |s_1 - s_2| \end{cases}$

Now, we may define prob. of events defined in terms of X and Y:

$P((X,Y) = (5,3)) := P(\{(s_1,s_2) \mid s_1 + s_2 = 5, |s_1 - s_2| = 3\})$

$= P(\{(4,1), (1,4)\})$

$= P(\{(4,1)\}) + P(\{(1,4)\})$

assuming independence

$= P(s_1 = 4)P(s_2 = 1) + P(s_1 = 1)P(s_2 = 4)$

assuming fair dice

$= (1/6)(1/6) + (1/6)(1/6) = 1/18$.

<u>Def</u>: Let $(X,Y)$ be a discrete bivariate vector.

Then the function from $\mathbb{R}^2 \to \mathbb{R}$:

$f(x,y) := P(X=x, Y=y)$ is the

<u>joint prob. mass function (pmf)</u> of $(X,Y)$

Remark: Denote by $f_{XY}(x,y)$ or $f_{X,Y}(x,y)$

where $f_{XY}(x,y)$ not necessarily $f_{YX}(x,y)$

Corollary: For $A \subset \mathbb{R}^2$, by AOP, we have that $P((X,Y) \in A) = \sum_{(x,y) \in A} f(x,y)$

$$\boxed{\begin{array}{l}\text{Prop. Let } g(x,y) \text{ be a function from } \mathbb{R}^2 \to \mathbb{R}. \\ \text{then } g(X,Y) \text{ is a real-valued RV: } \boxed{Eg(X,Y) := \sum_{(x,y) \in \mathbb{R}^2} g(x,y) f_{XY}(x,y)}\end{array}}$$

Prop (Linearity of Expectation): If $g_1, g_2$ are real-valued functions on $\mathbb{R}^2$; $a, b \in \mathbb{R}$, then

$$\boxed{E[a g_1(X,Y) + b g_2(X,Y)] = a E g_1(X,Y) + b E g_2(X,Y)}$$

Ex: Recall bivariate vector $(X,Y)$: $\begin{cases} X: \text{sum of the two dice} \\ Y: |\text{difference of two dice}| \end{cases}$

Let $f_{XY}(x,y)$ be joint pmf for $(X,Y)$

$P(X=Y) = P((X,Y) \in (\{1,1\}, \{2,2\}, \dots, \{6,6\}))$

$= \sum_{i=1}^{6} f_{XY}(i,i)$

$P(X=2)$ . . .

Note: $\mathcal{X} = \{2, \dots, 12\}$
$\mathcal{Y} = \{0, \dots, 5\}$

$P(X=x) = \sum_{y=0}^{5} f_{XY}(x,y)$

$P(Y=y) = \sum_{x=2}^{12} f_{XY}(x,y)$

Thm: Let $(X,Y)$ be a discrete bivariate RV w/ joint pmf $f_{XY}(x,y)$.

Then,
X and Y are discrete RV's w/ the following pmf's:

$$\boxed{\text{marginal pmf} := \begin{cases} f_X(x) = \sum_y f_{XY}(x,y) \\ f_Y(y) = \sum_x f_{XY}(x,y) \end{cases}}$$

Corollary: Joint Dist/pmf completely determines its marginal dist/pmf
But marginal Dist/pmf's do not determine the joint dist.

Ex: Let $X \sim \text{Bernoulli}(1/2)$
$Y \sim \text{Bernoulli}(1/2)$, $Y \perp\!\!\!\perp X$
$Z = 1 - X$

$\Rightarrow X \overset{d}{=} Y \overset{d}{=} Z$ ← why?

But $\underbrace{(X,Y)}_{X \perp\!\!\!\perp Y} \neq \underbrace{(X,Z)}_{X \not\perp\!\!\!\perp Z}$ in joint dist.

⋮

Show more

- Continuous Bivariate RV's are described via joint prob. density functions

Def: a function $f(x,y)$ from $\mathbb{R}^2 \to \mathbb{R}$ is called a joint pdf of the continuous bivariate random vector $(X,Y)$ if, for every $A \subseteq \mathbb{R}^2$

$$\boxed{P((X,Y) \in A) = \iint_{(x,y) \in A} f_{XY}(x,y)\,dx\,dy}$$

- $f(x,y) \geq 0 \;\forall (x,y) \in \mathbb{R}^2$
- $\iint_{\mathbb{R}^2} f(x,y)\,dx\,dy = 1$

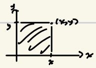
$E[g(X,Y)] := \iint_{\mathbb{R}} g(x,y) f_{XY}(x,y)\,dx\,dy$

<u>marginal pdfs</u> $:= f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$

$f_Y(y) = \int_{\mathbb{R}} f_{XY}(x,y)\,dx$

<u>Joint cdf</u> $:= F_{XY}(x,y) := P(X \le x, Y \le y)$

$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(s,t)\,dt\,ds$




- $0 \le F_X(x,y) \le 1 \quad \forall (x,y) \in \mathbb{R}^2$
- $F_{XY}(x,\cdot) \uparrow \quad \forall x \in \mathbb{R}$
- $F_{XY}(\cdot,y) \uparrow \quad \forall y \in \mathbb{R}$
- Fundamental thm. of calculus (bivariate case) gives

$$\frac{d^2 F_{XY}(x,y)}{dx\,dy} = f_{XY}(x,y)$$

at continuous points $(x,y)$ of function $f_{XY}$

## 4.2a: conditional distributions

Given a bivariate RV $(X,Y) \in \mathbb{R}^2$, we are interested in, say

$P(Y \in B \mid X \in A)$ for $A, B \subset \mathbb{R}$.

<u>Ex (conditional prob.)</u>: Let $(X,Y)$ = (height, weight) of a (random) person.

Find $P(Y > 200\text{ lb} \mid X = 6')$

2) $X$ = number people out of 100 sampled who vote for Smith

$Y$ = Smith wins

e.g. $P(Y = \text{Yes} \mid X = 53)$

<u>Prop.</u> Let $(X,Y)$ be discrete.

$$P(Y \in B \mid X \in A) = \frac{P(X \in A \cap X \in B)}{P(X \in A)}$$

$$P(X \in A \cap Y \in B) = \sum_{x \in A} \sum_{y \in B} f_{XY}(x,y)$$

$$P(X \in A) = \sum_{x \in A} \sum_{y} f_{XY}(x,y) \quad \text{, marginalizing on } Y$$

$$= \sum_{x \in A} f_X(x)$$

So,

$$P(Y \in B \mid X \in A) = \sum_{x \in A} \sum_{x \in B} f_{XY}(x,y) \Big/ \sum_{x \in A} f_X(x)$$

$$= \sum_{y \in B} \frac{\sum_{x \in A} f_{XY}(x,y)}{\sum_{x \in A} f_X(x)}$$

In particular, let $A = \{x\}$ and $B = \{y\}$. Then

$$P(Y = y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)} =: \text{conditional pmf}$$

$$\boxed{\text{Def: } f(y|x) = P(Y = y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\sum_y f_{XY}(x,y)}}$$

- $f(y|x) \ge 0 \quad \forall y$
- $\sum_y f(y|x) = \sum_y \frac{f_{XY}(x,y)}{\underbrace{\sum_y f_{XY}(x,y)}_{\text{depends only on } x}} = \frac{\sum_y f_{XY}(x,y)}{\sum_y f_{XY}(x,y)} = 1.$
- Denote $f_{Y|X}(y|x)$

<u>Remark</u>: both the marginal pmf $(f_X, f_Y)$ and conditional pmf $(f_{Y|X}, f_{X|Y})$ are completely determined by joint pmf $f_{XY}$

<u>Corollary</u>: $f_{Y|X}(y|x) = f_{XY}(x,y) / f_X(x)$

$$\Rightarrow \underline{f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x)}$$

Similarly,

$$\underline{f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y)}$$

whenever $f_X(x) > 0$ and $f_Y(y) > 0$

Now consider continuous bivariate random vector $(X,Y) \in \mathbb{R}^2$

Heuristic but incorrect argument: $P(Y \in B \mid X=x) = \dfrac{P(Y\in B, X=x)}{P(X=x)}$ ← technically $=0$ always

where $P(Y\in B, X=x) = \int_B f_{XY}(x,y)\,dy$

$P(X=x)$ "=" $f_X(x) = \int_{\mathbb{R}} f_{XY}(x,y)\,dy$

Hence

$$P(Y\in B \mid X=x) = \frac{\int_B f_{XY}(x,y)\,dy}{f_X(x)} \leftarrow \text{doesn't depend on } y$$

$$= \int_B \underbrace{\frac{f_{XY}(x,y)}{f_X(x)}}_{\text{correct identity}}\,dy \quad (*)$$

Problems: $P(X=x)=0$ and

$P(Y\in B, X=x) \le P(X=x) = 0$

The correct treatment is to use limits:

$$\boxed{P(Y\in B \mid X=x) := \lim_{\varepsilon\to 0} P(Y\in B \mid X\in(x-\varepsilon, x+\varepsilon))}$$

provided that such a limit exists & is unique

To assess the RHS and its limit, we need to use the joint pdf:

$$P(Y\in B \mid X\in(x-\varepsilon,x+\varepsilon)) = \frac{P(Y\in B \cap X\in(x-\varepsilon,x+\varepsilon))}{P(X\in(x-\varepsilon,x+\varepsilon))}$$

$$= \frac{\int_B \int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx\,dy}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}$$

$$= \lim_{\varepsilon\to 0} \int_B \frac{\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx\,dy}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}\,dy$$

$$= \int_B \lim_{\varepsilon\to 0} \frac{\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f_{XY}(x,y)\,dx\,dy}{\frac{1}{2\varepsilon}\int_{x-\varepsilon}^{x+\varepsilon} f_X(x)\,dx}\,dy$$

where $2\varepsilon$: range of integral bounds

and by L'Hopital's Rule

$$= \int_B \frac{f_{XY}(x,y)}{f_X(x)}\,dy$$

Def: Let $(X,Y)$ be bivariate RV w/ joint pdf $f_{XY}(x,y)$. Then

$$\boxed{f_{Y|X}(y|x) := \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)\,dy}}$$

Remarks: $\cdot\ f_{Y|X}(x,y) \ge 0 \quad \forall y, x \ (f_X(x)>0)$

$\cdot \int f_{Y|X}(y|x)\,dy = 1$

$\cdot$ Identities hold: $f_{XY}(x,y) = \begin{cases} f_X(x) f_{Y|X}(y|x) \\ f_Y(y) f_{X|Y}(x,y) \end{cases}$

Notice: $f_{Y|X}(y|x) = \{f_{Y|X}(y|x)\}_{x\in\mathbb{R}}$

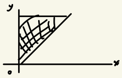
Remark: Suppose $f(\underset{\text{data points}}{x}|\theta)$. Then $\int f(x|\theta)\,dx = 1$ b/c $f_X(x)$ is a (valid) pdf

whereas $f(x|\theta)$ as a function of $\theta$ is defined as "likelihood"

Def (Cond. Expectation): Given $(X,Y)\sim f_{XY}$, $g:\mathbb{R}\to\mathbb{R}$ a function

then,
$$\boxed{\begin{aligned} E[g(Y)|x] &:= E[g(Y)|X=x] \\ &:= \int g(y) f_{Y|X}(y|x)\,dy \end{aligned}}$$

**Ex**: Let $(X,Y) \sim f_{XY}(x,y) = \begin{cases} e^{-y}, & 0<x<y<\infty \\ 0, & \text{otherwise} \end{cases}$



check $f_{XY}(x,y) \geq 0 \ \forall\, x,y \in \mathbb{R}^2$.

$$\int_{\mathbb{R}^2} f_{XY}(x,y)\,dx\,dy = \int_0^\infty \int_0^y e^{-y}\,dx\,dy = \int_0^\infty y e^{-y}\,dy$$

$$= \int_0^\infty (-y)\,de^{-y}$$

$$\overset{\text{IBP}}{=} \int_0^\infty e^{-y}\,dy = 1$$

$\Rightarrow$ valid joint pdf. ∎