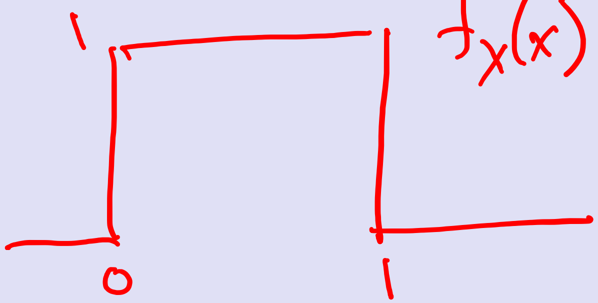
ENGG 2430 / ESTR 2004: Probability and Statistics

Spring 2019

8. Everything you wanted to know about random variables but were afraid to ask

Andrej Bogdanov

X **is** Uniform$(0, 1)$. **What is the PDF of** $Y = \sqrt{X}$**?**



$P(X \le x) = x$ WHEN $0 \le x \le 1$.

$P(Y \le y) = P(\sqrt{X} \le y) = P(X \le y^2) = y^2$ WHEN $0 \le y \le 1$

$2 \quad f_Y(y) = \frac{d}{dy} P(Y \le y) = 2y$

$V = \cancel{50}$ Exponential(1) km/h ←

~~50km~~ 1km

0km

What is the PDF of the travel time?

$T = \dfrac{\cancel{50}\,1}{V}$ V is Exponential(1)

$P(T \le t) = P\left(\frac{1}{V} \le t\right) = P\left(V \ge \frac{1}{t}\right) = e^{-1/t} \quad 0 \le t \le \infty$

$f_T(t) = \frac{d}{dt} P(T \le t) = \frac{1}{t^2} e^{-1/t} \quad 0 \le t \le \infty$

Shifting and scaling

$$Y = aX + b$$

$$P(Y \le y) = P(aX + b \le y)$$

$$= P\left(X \le \frac{y-b}{a}\right) \quad \text{IF } a > 0$$

$$f_Y(y) = \frac{d}{dy} P\left(X \le \frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)}$$

Normalization

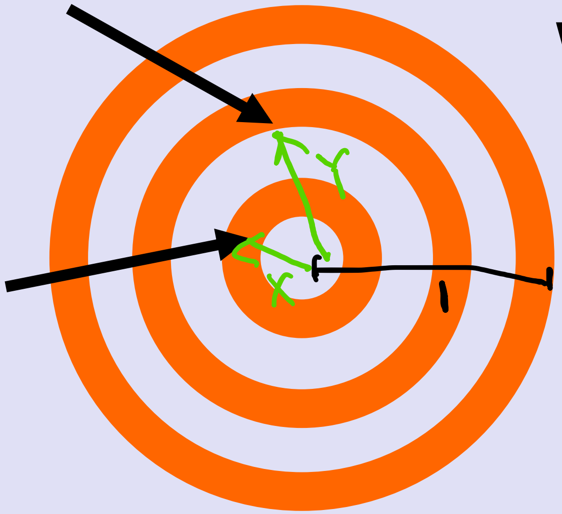
X **is** $Normal(\mu, \sigma)$

$Y = (X - \mu)/\sigma$

$E[Y] = (E[X] - \mu)/\sigma = 0$
$Var[Y] = Var[(X-\mu)/\sigma] = 1$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$f_Y(y) = \sigma f_X\left(\frac{y + \mu/\sigma}{1/\sigma}\right) = \sigma f_X(\sigma y + \mu)$$
$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$



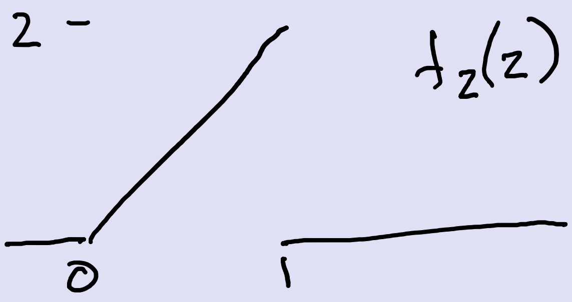
What is the PDF of the loser?

X, Y Uniform$(0,1)$ INDEPENDENT

$Z = \max\{X, Y\}$

$$P(Z \le z) = P(X \le z, Y \le z) = P(X \le z)P(Y \le z) = z^2$$

$$f_Z(z) = \frac{d}{dz} P(Z \le z) = 2z \qquad 0 \le z \le 1$$



X, Y **are independent** $\text{Uniform}(0, 1)$.

What is the PDF of Y/X**?**

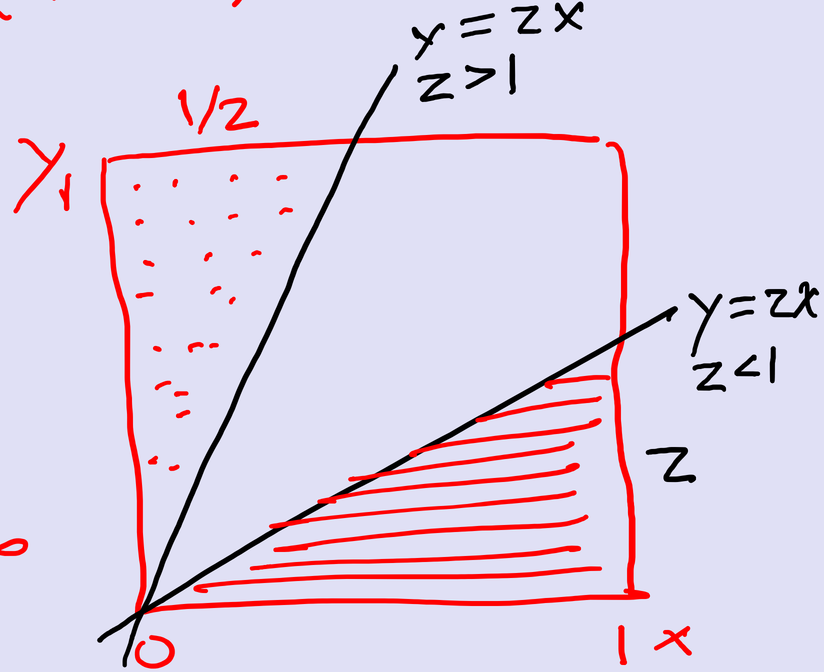
$Z = Y/X$

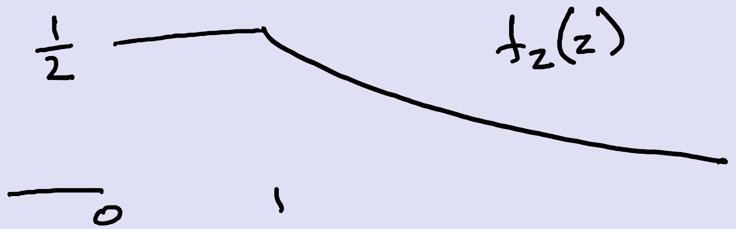
$P(Z \le z) = P(Y/X \le z) = P(Y \le zX)$

$z \le 1 \quad P(Y \le zX) = \frac{z}{2}$

$z > 1 \quad P(Y \le zX) = 1 - \frac{1}{2z}$

$$f_Z(z) = \begin{cases} 1/2 & \text{WHEN } 0 < z < 1 \\ 1/2z^2 & \text{WHEN } 1 < z < \infty \end{cases}$$





Convolution

X, Y **are independent.**

What is the PMF/PDF of $\underbrace{X+Y}_{Z}$**?**

$$P(Z=z) = \sum_{x,y:\,x+y=z} P(X=x, Y=y)$$

$$= \sum_{x,y:\,x+y=z} P(X=x)P(Y=y) \qquad y=z-x$$

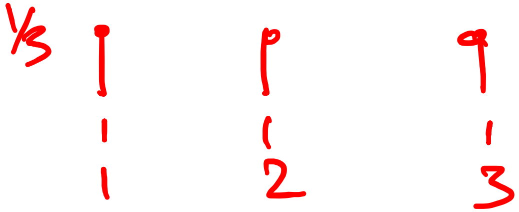
$$= \sum_{x} P(X=x)P(Y=z-x)$$

$$\boxed{f_Z(z) = \sum_x f_X(x) f_Y(z-x)} \quad \text{DISCRETE}$$

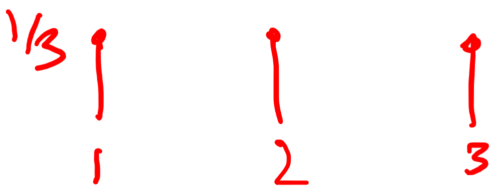
$$\boxed{f_Z(z) = \int_x f_X(x) f_Y(z-x)\,dx} \quad \text{CONTINUOUS}$$

X, Y are face values of 3-sided dice. What is the PMF of $X + Y$?

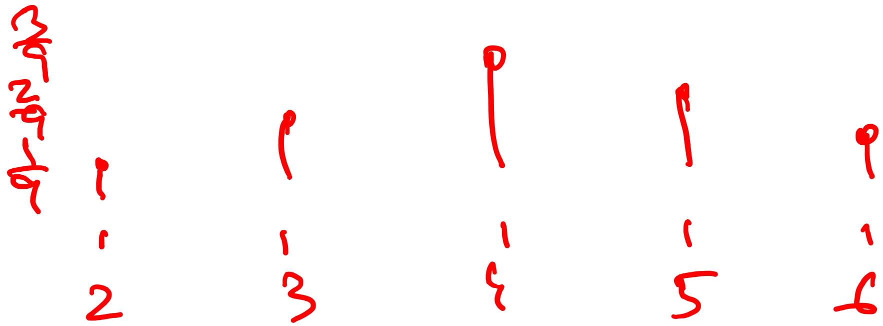
PMF OF X



PMF OF Y



PMF OF X+Y



Romeo and Juliet arrive in Shatin at independent $\mathrm{Exponential}(1)$ **hours past noon. How long is the wait?**

X = ROMEO   Y = JULIET   X, Y Exponential(1)

$Z = X - Y$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z+y)\, f_Y(y)\, dy$$

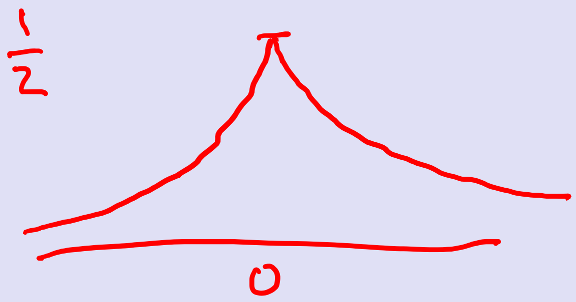
$z > 0$:

$$f_Z(z) = \int_0^{\infty} e^{-(z+y)} e^{-y}\, dy = e^{-z} \int_0^{\infty} e^{-2y}\, dy$$

$$= \frac{1}{2} e^{-z}$$

$z < 0$: $f_Z(z) = \frac{1}{2} e^{z}$

$$\boxed{f_Z(z) = \frac{1}{2} e^{-|z|}.}$$



Sum of Independent Normals

X, Y Normal$(0,1)$ $\quad f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$Z = X + Y$

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \cdot e^{-\frac{(z-x)^2}{2}} dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2 + xz - z^2/2} dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(x - z/2)^2} e^{-z^2/4} dx$$

$$= \frac{1}{2\pi} e^{-z^2/4} \underbrace{\int_{-\infty}^{\infty} e^{-(x - z/2)^2} dx}$$

INDEPENDENT OF Z

$$= \frac{C}{2\pi} e^{-z^2/4} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{2}} e^{-z^2/4} \quad \text{Normal}(0, \sqrt{2})$$

Sum of Independent Normals

$X = \text{Normal}(\mu, \sigma)$

$Y = \text{Normal}(\mu', \sigma')$

INDEPENDENT

$X+Y = \text{Normal}(\mu+\mu', \sqrt{\sigma^2+\sigma'^2})$

Measuring dependence

$$\mathbf{Cov}[X, Y] = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

$= E[XY] - E[X]E[Y]$

$\mathrm{Var}[X] = \mathrm{Cov}[X, X]$

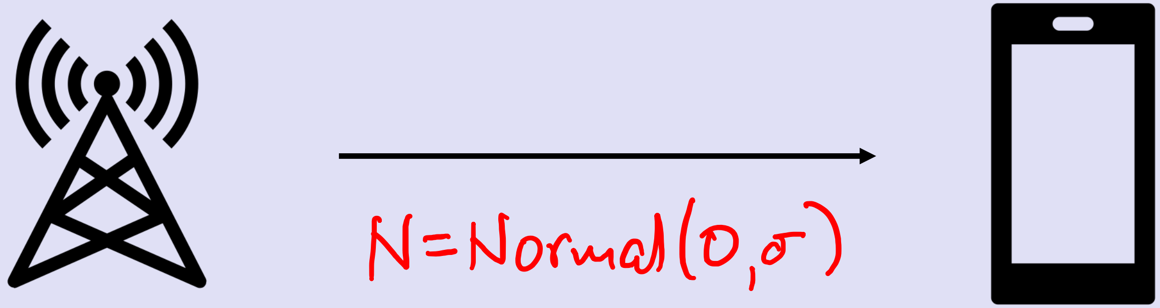
$$\mathbf{Corr}[X, Y] = \frac{\mathbf{Cov}[X, Y]}{\sqrt{\mathbf{Var}[X]\ \mathbf{Var}[Y]}}$$

ALWAYS BETWEEN -1 AND 1

Measuring dependence

	Cov	Corr
X, Y **ind.**	0	0
$Y = X$	Var[X]	1
$Y = -X$	-Var[X]	-1

Warning: Zero-covariance is not independence!



$N = \text{Normal}(0, \sigma)$

$X = 1$ **or** -1

$\quad\ \tfrac{1}{2} \qquad \tfrac{1}{2}$

$Y = X + N$

$$\textbf{Cov}[X, Y] = E[XY] - \overset{0}{E[X]}E[Y] = E[X(X+N)]$$

$$= E[X^2] + E[XN] \qquad E[XN] = E[X]E[N] = 0$$

$$= 1$$

$$\textbf{Corr}[X, Y] = \frac{Cov[X,Y]}{\sqrt{\underbrace{Var[X]}_{=1}Var[Y]}} = \frac{1}{\sqrt{Var[X+N]}} = \frac{1}{\sqrt{Var[X]+Var[N]}}$$

$$= \frac{1}{\sqrt{1+\sigma^2}}$$

Variance of a sum

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$$ **if independent**

$$
\begin{aligned}
\mathrm{Var}[X+Y] &= E\left[\left(X+Y-E[X+Y]\right)^2\right] \\
&= E\left[\left((X-E[X]) + (Y-E[Y])\right)^2\right] \\
&= E\left[(X-E[X])^2\right] + E\left[(Y-E[Y])^2\right] \\
&\quad + 2E\left[(X-E[X])(Y-E[Y])\right] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X,Y] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + \mathrm{Cov}[X,Y] + \mathrm{Cov}[Y,X]
\end{aligned}
$$

Variance of a sum

$$\mathbf{Var}[X_1 + \ldots + X_n] = \mathbf{Var}[X_1] + \ldots + \mathbf{Var}[X_n]$$

if every pair X_i, X_j is independent.

IN GENERAL

$$\mathrm{Var}[X_1 + \ldots + X_n] = \underbrace{\sum_{i=1}^{n} \mathrm{Var}[X_i]}_{n \text{ TERMS}} + \sum_{i \neq j} \underbrace{\mathrm{Cov}[X_i, X_j]}_{n(n-1) \text{ TERMS}}$$

n people throw their hats in a box and pick one out at random. What is the variance of the number of people who get their own?

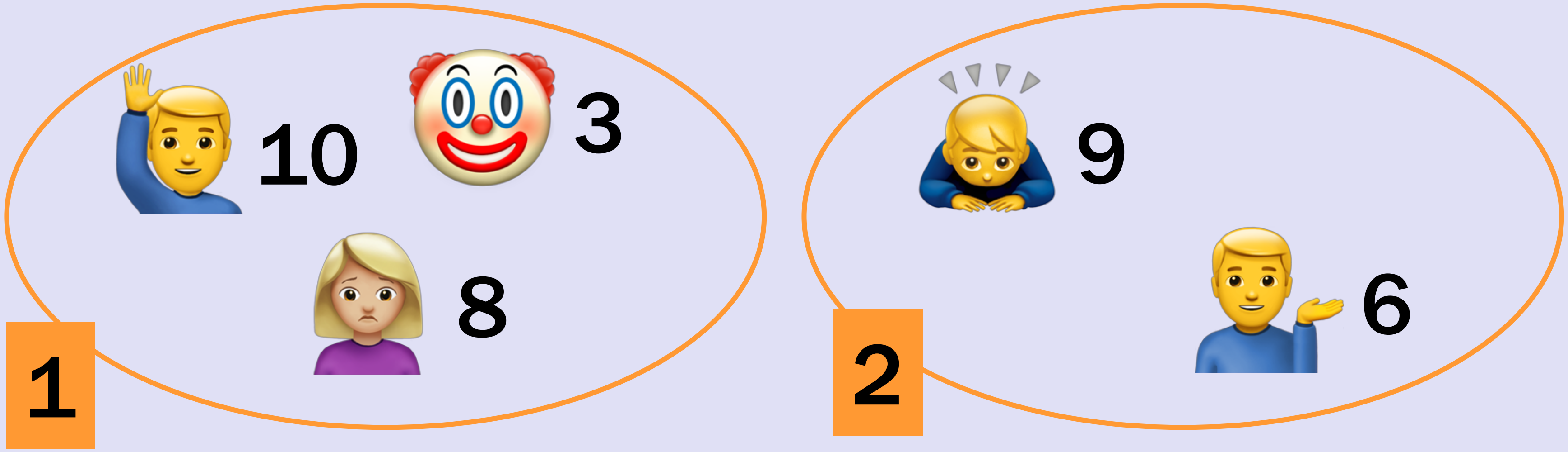
$X = X_1 + X_2 + \dots + X_n \qquad X_i = \begin{cases} 1 & \text{IF } i\text{TH P.} \\ 0 & \text{IF NOT} \end{cases}$

$E[X] = E[X_1] + \dots + E[X_n] = n \cdot \frac{1}{n} = 1$

$\mathrm{Var}[X] = \underbrace{\mathrm{Var}[X_1]}_{\frac{1}{n}\left(1-\frac{1}{n}\right)} + \dots + \underbrace{\mathrm{Var}[X_n]}_{\frac{1}{n}\left(1-\frac{1}{n}\right)} + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j]$

$\mathrm{Cov}[X_i, X_j] = \underbrace{E[X_i X_j]}_{1/n \cdot 1/(n-1)} - \underbrace{E[X_i]E[X_j]}_{1/n^2} = \frac{1}{n^2(n-1)}$

$\mathrm{Var}[X] = n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) + n(n-1) \cdot \frac{1}{n^2(n-1)} = 1$



$X =$ **score**

$Y =$ **section**

$\mathbf{E}[X \mid Y]$:

7	7.5
3/5	2/5
Y=1	Y=2

$$\mathbf{E}[X] = 7 \cdot \frac{3}{5} + 7.5 \cdot \frac{2}{5} = 7.2$$

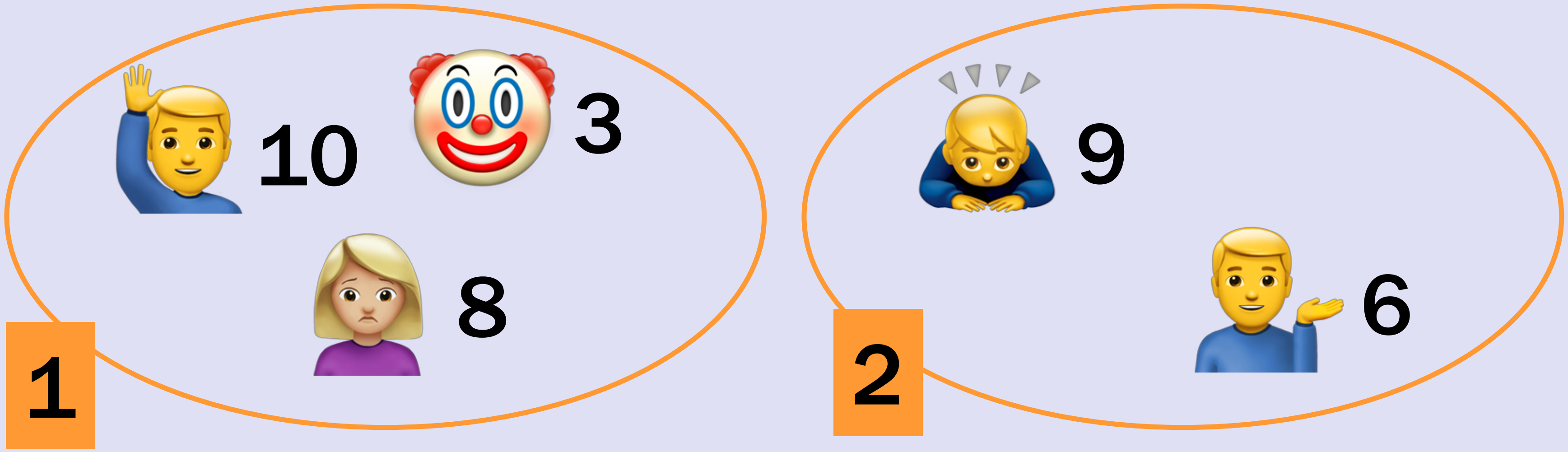
Estimation

$\hat{X} = \mathbf{E}[X \mid Y]$ is an estimator for X given Y

$$\boxed{\mathbf{E}[X] = \mathbf{E}[\hat{X}] = \mathbf{E}[\mathbf{E}[X \mid Y]]}$$

Total expectation theorem

$\tilde{X} = X - \hat{X}$ is the error



$\hat{X} = \mathbf{E}[X \mid Y]$:	7	7	7	7.5	7.5
$\tilde{X} = X - \hat{X}$:	3	1	-4	1.5	-1.5
$\mathbf{E}[\tilde{X} \mid Y]$:		0		0	
$\mathbf{E}[\hat{X}\tilde{X} \mid Y]$:		0		0	

Estimation

$\hat{X} = \mathbf{E}[X \mid Y]$ **is an estimator for** X **given** Y

$\tilde{X} = X - \hat{X}$ **is the error**

$\mathbf{E}[\tilde{X} \mid Y] = 0$

$\mathbf{E}[\hat{X}\tilde{X} \mid Y] = 0$

$\mathbf{Cov}[\hat{X}, \tilde{X}] = 0$

Conditional variance

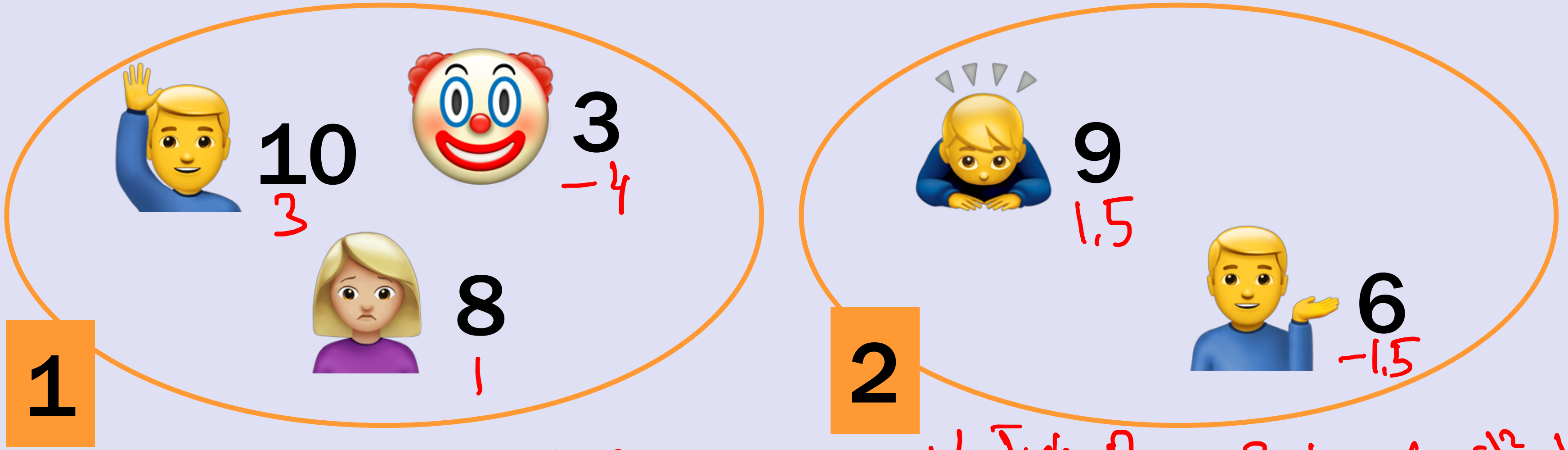
$$\mathbf{Var}[X \mid Y] = \mathbf{E}[\widetilde{X}^2 \mid Y]$$

$X = \hat{X} + \breve{X}$

Total variance theorem:

$$\begin{aligned}\mathbf{Var}[X] &= \mathbf{Var}[\hat{X}] + \mathbf{Var}[\widetilde{X}] \\ &= \underbrace{\mathbf{Var}[\mathbf{E}[X \mid Y]]}_{\text{VARIANCE BETWEEN GROUPS}} + \underbrace{\mathbf{E}[\mathbf{Var}[X \mid Y]]}_{\text{VARIANCE WITHIN EACH GROUP}}\end{aligned}$$

$E[X|Y=1]=7$ $E[X|Y=2]=7.5$



$$\mathrm{Var}[X|Y=1] = 3^2\cdot\frac{1}{3} + (-4)^2\cdot\frac{1}{3} + 1^2\cdot\frac{1}{3} = \frac{26}{3}$$

$$\mathrm{Var}[X|Y=2] = 1.5^2\cdot\frac{1}{2} + (-1.5)^2\cdot\frac{1}{2}$$

$\hat{X} = \mathbf{E}[X \mid Y]$:

7	7.5
3/5	2/5

$E[X]=7.2$

$$\mathbf{Var}\ \mathbf{E}[X \mid Y] = (-0.2)^2\cdot\frac{3}{5} + (0.3)^2\cdot\frac{2}{5} = 0.06$$

$\mathbf{Var}[X \mid Y]$:

26/3	9/4
3/5	2/5

$$\mathbf{E}\ \mathbf{Var}[X \mid Y] = \frac{26}{3}\cdot\frac{3}{5} + \frac{9}{4}\cdot\frac{2}{5} = 6.1 \longrightarrow \boxed{\mathrm{Var}[X]=6.16}$$

P **is** Uniform(0, 1), X **is** Binomial(n, P)

What is Var[X]?

$$E[X] = E[E[X|P]] = E[nP] = nE[P] = \frac{n}{2}$$

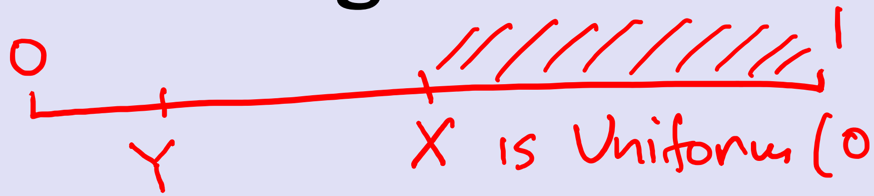
$$\mathrm{Var}[X] = \mathrm{Var}[E[X|P]] + E[\mathrm{Var}[X|P]]$$

$$= \mathrm{Var}[nP] + E[nP(1-P)]$$

$$= n^2 \underbrace{\mathrm{Var}[P]}_{1/12} + n(\underbrace{E[P]}_{1/2} - \underbrace{E[P^2]}_{1/3})$$

$$= \frac{n^2}{12} + \frac{n}{6}$$

Break a stick of length 1 at a random point. Keep the left part and repeat. What is the $\mathbf{E}$ and $\mathbf{Var}$ of the length?



X is Uniform$(0,1)$

Y is Uniform$(0,X)$

$$E[Y] = E[E[Y|X]] = E\left[\frac{X}{2}\right] = \frac{1}{2}E[X] = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

$$\begin{aligned}
Var[Y] &= Var[E[Y|X]] + E[Var[Y|X]] \\
&= Var\left[\frac{X}{2}\right] + E\left[\frac{X^2}{12}\right] \\
&= \frac{1}{4}Var[X] + \frac{1}{12}E[X^2] \\
&= \frac{1}{4}\cdot\frac{1}{12} + \frac{1}{12}\cdot\frac{1}{3} = \frac{1}{48} + \frac{1}{36}
\end{aligned}$$