More Random Variables

Geometric (p) – represents repeated coin flipping until 1st head

PMF: $P[X=i] = (1-p)^{i-1} p$	$P[X>i] = (1-p)^i$
$E[X] = \sum_{i=1}^{\infty} i(1-p)^{i-1} p = p \frac{d}{dp}\left[-\sum_{i=1}^{\infty}(1-p)^i\right] = -p \frac{d}{dp}\left[\frac{1}{p} - 1\right] = -p\left(-\frac{1}{p^2}\right) = \boxed{\frac{1}{p}}$	$Var(X) = \boxed{\frac{1-p}{p^2}}$

Poisson (λ) – represents an "arrival process"

PMF: $P[X=i] = \frac{\lambda^i}{i!} e^{-\lambda}$	
$E[X] = \boxed{\lambda}$	$Var(X) = \boxed{\lambda}$

Covariance: measures association between R.V.'s

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

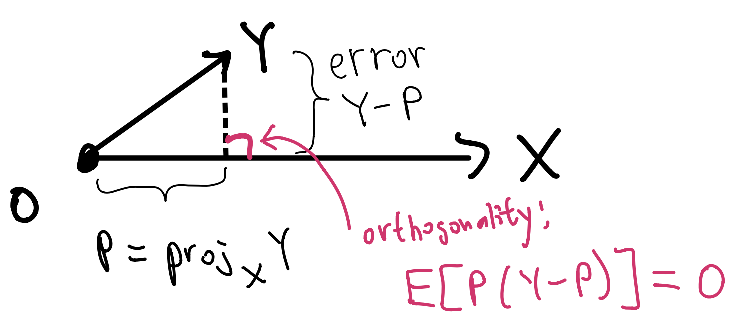
$$\boxed{Cov(X,X) = Var(X)}$$

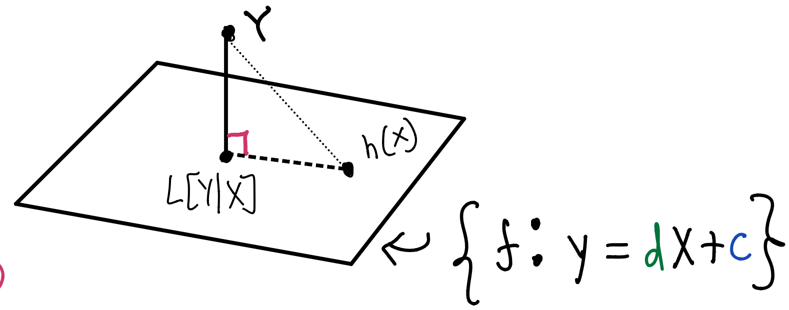
Bilinearity

$$\boxed{Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j Cov(X_i, Y_j)}$$

LLSE

$$\boxed{L[Y|X] = E[Y] + \frac{cov(X,Y)}{Var(X)}(X - E[X])}$$





1 Geometric and Poisson

Let $X \sim \text{Geo}(p)$ and $Y \sim \text{Poisson}(\lambda)$ be independent. random variables. Compute $\mathbb{P}(X > Y)$. Your final answer should not have summations.

$$P(X>Y) = \sum_{i=0}^{\infty} P(X>Y \mid Y=i)\, P(Y=i)$$

$$= \sum_{i=0}^{\infty} P(X>i)\, P(Y=i)$$

$$= \sum_{i=0}^{\infty} (1-p)^i\, e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

$$= \frac{1}{e^{\lambda p}} \sum_{i=0}^{\infty} \frac{(\lambda-\lambda p)^i}{i!}\, e^{-\lambda}\, e^{\lambda p}$$

$$= \frac{1}{e^{\lambda p}} \sum_{i=0}^{\infty} \frac{(\lambda-\lambda p)^i}{i!}\, e^{-(\lambda-\lambda p)}$$

$Z \sim \text{Poisson}(\lambda-\lambda p)$

$$= \frac{1}{e^{\lambda p}}$$

Confidence Intervals

Apply concentration inequalities to set up bounds and solve for desired number!

2 Vegas

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip his or her coin. Assume that each person is independently likely to carry a fair or a trick coin.

$\frac{1}{2}p+\frac{1}{2}$

$$X_i = \begin{cases} 1 & \text{w.p. } p \cdot 1 + (1-p)\cdot\frac{1}{2} \\ 0 & \text{o/w} \end{cases}$$

1. Given the results of your experiment, how should you estimate p? (*Hint:* Construct an (unbiased) estimator for p such that $E[\hat{p}] = p$.)

$$Z = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad E[Z] = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$= \cancel{\frac{1}{n}}\cancel{n}\left(\frac{1}{2}p+\frac{1}{2}\right) = \frac{1}{2}p+\frac{1}{2}$$

$$p = 2E[Z]-1 = E[2Z-1] \qquad \boxed{\hat{p} = 2Z-1}$$

2. How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

$$E[\hat{p}] = p$$

$$Var(\hat{p}) = Var\left(2\cdot\frac{1}{n}\sum_{i=1}^{n}X_i - 1\right)$$

$$= \frac{4}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{4}{n^{2}}\cancel{n}\cdot q(1-q)$$

$$P[|\hat{p}-p| \le 0.05] \ge 0.95$$

$$1 - P[|\hat{p}-p| \le 0.05] \le 1-0.95$$

$$P[|\hat{p}-p| > 0.05] \le 0.05$$

$$P[|\hat{p}-p|>0.05] \le P[|\hat{p}-p| \ge 0.05] \le \frac{\frac{4}{n}q(1-q)}{0.05^2} \le \frac{\frac{4}{n}\cdot\frac{1}{2}\cdot\frac{1}{2}}{0.05^2} \le \frac{1}{0.05^2 n} \le 0.05$$

$$n \ge \frac{1}{0.05^3} = \boxed{8000}$$

3 LLSE

We have two bags of balls. The fractions of red balls and blue balls in bag A are $2/3$ and $1/3$ respectively. The fractions of red balls and blue balls in bag B are $1/2$ and $1/2$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. You then draw 6 balls from that same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1\le i\le 3} X_i$ and $Y = \sum_{4\le i\le 6} X_i$. Find $L(Y \mid X)$. *Hint*: Recall that

$$L(Y \mid X) = \mathbb{E}(Y) + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var}(X)}\big(X - \mathbb{E}(X)\big).$$

Also remember that covariance is bilinear.

$$X_i = \begin{cases} 1 & \text{wp } \frac{1}{2}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{2} = 7/12 \\ 0 & \text{wp } 5/12 \end{cases}$$

$$X_i^2 = X_i \qquad X_iX_j = \begin{cases} 1 & \text{wp } \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{2}{3} + \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{25}{72} \\ 0 & \text{wp } \frac{47}{72} \end{cases}$$

$$E[X_i^2] = \frac{7}{12}$$

$$E[X_i] = \frac{7}{12} \qquad E[X_iX_j] = \frac{25}{72}$$

$$\operatorname{Var}(X_i) = \frac{7}{12} - \left(\frac{7}{12}\right)^2 = \frac{35}{144}$$

$$\operatorname{Cov}(X_i, X_j) = E[X_iX_j] - E[X_i]E[X_j] = \frac{25}{72} - \left(\frac{7}{12}\right)^2 = \frac{1}{144}$$

$$E[X] = E[Y] = 3E[X_i] = \frac{7}{4}$$

$$\operatorname{cov}(X,Y) = \operatorname{cov}\left(\sum_{i=1}^{3} X_i, \sum_{j=4}^{6} X_j\right) = \sum_{i=1}^{3}\sum_{j=4}^{6} \operatorname{cov}(X_i X_j) = \sum_{i=1}^{3}\sum_{j=4}^{6} \frac{1}{144}$$

$$9\operatorname{cov}(X_iX_j) = \frac{1}{16}$$

$$\operatorname{Var}(X) = \operatorname{cov}\left(\sum_{i=1}^{3} X_i, \sum_{j=1}^{3} X_j\right) = \sum_{i=1}^{3}\sum_{j=1}^{3} \operatorname{cov}(X_i, X_j)$$

$$= 3\operatorname{var}(X_i) + 6\operatorname{cov}(X_i, X_j) = 3\cdot\frac{35}{144} + 6\,\frac{1}{144} = \frac{37}{48}$$

i	1	1	1	2	2	2	3	3	3
j	1	2	3	1	2	3	1	2	3

$$L[Y|X] = E[Y] + \frac{\operatorname{cov}(X,Y)}{\operatorname{Var}(X)}\left(X - E[X]\right)$$

$$= \frac{7}{4} + \frac{1/\cancel{16}}{37/\cancel{48}_{3}}\left(X - \frac{7}{4}\right)$$

$$= \frac{7}{4} + \frac{3}{37}X - \frac{21}{148}$$

$$= \boxed{\frac{3}{37}X + \frac{119}{74}}$$