Lecture 10

8.2

Expectation and Independence

① If $X_1, \dots, X_n$ are n independent RV's

then:

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E(X_i)$$

② Let $g_1, \dots, g_n$ be $\mathbb{R} \to \mathbb{R}$ functions, then

along with ①

$$E\left[\prod_{i=1}^{n} g(X_i)\right] = \prod_{i=1}^{n} E(g(X_i))$$

Variance

- If X is a RV then:

$$Var(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

- If $X_1, \dots, X_n$ are n independent RV's then:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Ex) The plumber will come to your apartment at any time between 1pm - 7pm.

Once they have arrived, the time it takes for them to fix the sink is exponentially distributed with mean 30 min.

1. What is the expected time the plumbes will have finished the sink?

2. What is the variance of this time (in hours).

① Let X denote the arrival time

Let Y denote the time to fix

So the time the sink is finished,

$T = X + Y$.

$X \sim \text{Unif}(1, 7)$

$Y \sim \exp\left(\frac{1}{.5}\right) = \exp(2)$ $\left(\begin{array}{c} E(Y) = \frac{1}{\lambda} = .5 \\ \Rightarrow \lambda = 2 \end{array}\right)$

$E(T) = E(X) + E(Y) = \frac{7+1}{2} + .5 = 4.5$

$= 4{:}30$ PM

② $Var(T) = Var(X+Y) = Var(X) + Var(Y)$

$Var(X) = \frac{(b-a)^2}{12} = \frac{(7-1)^2}{12} = \frac{36}{12} = 3$

$Var(Y) = \frac{1}{\lambda^2} = \frac{1}{2^2} = \frac{1}{4}$

$\Rightarrow Var(T) = 3\frac{1}{4}$

Coupon Collector's Problem

Ash Ketchum is trying to complete his pokedex of n pokemon. Everytime Ash sees a pokemon he uses a pokeball to catch it, and it is added to his pokedex. Every pokemon has a $\frac{1}{n}$ chance of appearing.

1. What is the expected number of pokeballs required for Ash to complete his pokedex?

2. What is the variance?

① Let X denote the number of pokeballs Ash uses.

$X = T_1 + \dots + T_n$ where T_i is the number of tosses required to catch your i'th unique pokemon after your $i-1$'st

So $T_i \sim \text{geom}(p_i)$.

$p_1 = 1$

$p_2 = \frac{n-1}{n}$

$\vdots$

$p_n = \frac{n-(n-1)}{n} = \frac{1}{n}$

So $E(X) = E(T_1 + \dots + T_n)$

$$= 1 + \frac{n}{n-1} + \dots + \frac{n}{1}$$

$$= n\left(\sum_{k=1}^{n} \frac{1}{k}\right).$$

(2) $Var(X) = Var(T_1 + \ldots + T_n)$

$= Var(T_1) + \ldots + Var(T_n)$

If $T_i \sim geom(p_i) \Rightarrow Var(T_i) = \dfrac{1-p_i}{p_i^2}$

$p_i = \dfrac{n-i+1}{n}$

$\Rightarrow Var(X) = \sum_{i=1}^{n} \left(\dfrac{1-p_i}{p_i^2} \right)$

Empirical/Sample Mean and Variance

Let $X_1, \dots, X_n$ be n i.i.d. RV with mean μ and variance σ^2.

- Sample Mean $:= \frac{\sum X_i}{n} = \overline{X}_n$

① What is $Var(\overline{X}_n)$?

$$Var(\overline{X}_n) = Var\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} Var(\sum X_i)$$

$$= \frac{1}{n^2} \cdot \sum Var(X_i)$$

$$= \frac{n}{n^2} Var(X_1)$$

$$= \frac{\sigma^2}{n}.$$

Unbiased Estimators

Let θ be the parameter of a distribution for RV X (ex: $\theta = Var(X)$ or mean, etc)

Define an estimator for θ, $\hat{\theta}_n : \mathbb{R}^n \to \mathbb{R}$

1. $Bias(\hat{\theta}_n) := E(\hat{\theta}_n) - \theta$

2. $\hat{\theta}_n$ is unbiased if bias $= 0$, i.e. $E(\hat{\theta}_n) = \theta$

Ex)

Let $X_1, \dots, X_n$ be n independent RV s.t. $Var(X) = \sigma^2$
$E(X) = \mu$

Find c such that $E\left(c \cdot \sum_{i=1}^{n} (X_i - \overline{X}_n)^2\right) = \sigma^2$

$$E\left(\sum (X_i - \overline{X}_n)^2\right) = \sum E\left[(X_i - \overline{X}_n)^2\right]$$

$$= \sum E\left[(X_i - \overline{X}_n)^2\right]$$

$$= \sum E\left[(X_i - \mu + \mu - \overline{X}_n)^2\right]$$

$$= \sum E\left[(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \overline{X}_n) + (\mu - \overline{X}_n)^2\right]$$

$$= \sum E\left[(X - \mu)^2\right] + 2\sum E\left[(X_i - \mu)(\mu - \overline{X}_n)\right] + \sum E(\mu - \overline{X}_n)^2$$

$$= n \cdot \sigma^2 + \underbrace{2\sum E\left[(X_i - \mu)(\mu - \overline{X}_n)\right]} + \sum E(\mu - \overline{X}_n)^2$$

$$E\left[\sum (X_i - \mu)(\mu - \overline{X}_n)\right]$$

$$= E\left[(\mu - \overline{X}_n) \cdot \left(\sum X_i - n\mu\right)\right]$$

$$= E\left[(\mu - \overline{X}_n)(n\overline{X}_n - n\mu)\right]$$

$$= -n \cdot E\left((\mu - \overline{X}_n)^2\right)$$

$$= n\sigma^2 - 2n E\left[(\mu - \overline{X}_n)^2\right] + n E\left[(\mu - \overline{X}_n)^2\right]$$

$$= n\sigma^2 - n \cdot \left(\frac{\sigma^2}{n}\right) = (n-1)\sigma^2$$

So $c = \frac{1}{n-1}$ s.t. $E\left(\frac{1}{n-1} \sum (X_i - \overline{X}_n)^2 \right) = \sigma^2$