6. Let X denote the number of Canon SLR cameras sold during a particular week by a certain store. The pmf of X is

x	0	1	2	3	4
$p_X(x)$	.1	.2	.3	.25	.15

Sixty percent of all customers who purchase these cameras also buy an extended warranty. Let Y denote the number of purchasers during this week who buy an extended warranty.

a. What is $P(X = 4, Y = 2)$? [*Hint*: This probability equals $P(Y = 2 \mid X = 4) \cdot P(X = 4)$; now think of the four purchases as four trials of a binomial

$P(Y=y \mid X=x)$.

$P(X=4, Y=2) = P($ You sold four camera sales and 2 sales also bought the extended warranty $)$

$$= P(X=4 \mid Y=2) \cdot P(Y=2) \quad \text{or} \quad P(Y=2 \mid X=4) \cdot P(X=4)$$

$P(Y=2 \mid X=4)$ = probability of getting 2 two extended warranties in 4 sales.

$P_{Y|X}(y \mid x=4)$

~ can be modelled by a binomial dist with $n=4$ and $p=0.6$

$B \sim Bin(n=4, p=0.6)$

$$= P(B=2)$$

$$= \binom{4}{2} \cdot (0.6)^2 \cdot (0.4)^2$$

$$\therefore P(X=4, Y=2) = P(Y=2 \mid X=4) \cdot P(X=4)$$

$$= \left(\frac{4 \times 2}{2} \cdot (0.6)^2 \cdot (0.4)^2 \right) \cdot (0.15)$$

Poisson Process

need the rate α to define such a process.

$N(t)$ = # occurences in the time int t.

$P(N(t) = k)$ = Pois with parameter $\lambda = \alpha t$.

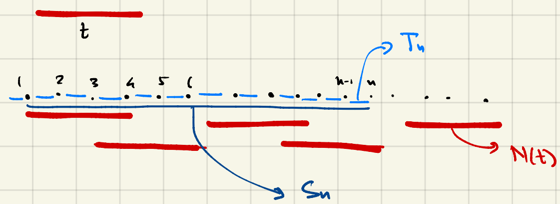
$$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

$$e^{\lambda} = \sum_{i=0}^{\infty} \left(\frac{\lambda^x}{x!} \right)$$

$$1 = \sum_{i=0}^{\infty} \left(\frac{\lambda^x}{x!} \right) \cdot e^{-\lambda}$$

$\lambda = \alpha t$

$$P(N(t) = k) = e^{-\alpha t} \cdot \frac{(\alpha t)^k}{k!}$$



T_n = time until the n event happens after $(n-1)^{th}$ event

= time betwn two successive events.

$T_n \sim exp(\alpha)$

S_n = amount of time until the n^{th} occurence.

$$= \sum_{i=1}^{n} T_n.$$

$T_i \sim exp(\lambda)$, $i = 1, 2, \ldots, n$

What is the distribution of S_n?

Note $exp(\lambda) = Gamma(\alpha = 1, \beta = \frac{1}{\lambda})$

$\because T_i \sim exp(\lambda) = Gam(\alpha = 1, \beta = \frac{1}{\lambda})$

$$S_n = Gamma(\alpha = n, \beta = \frac{1}{\lambda})$$

$N(t)$ = # of arrivals/occurences in time int of length t

T_n = time betwn $(n-1)^{th}$ and n^{th} occurence.

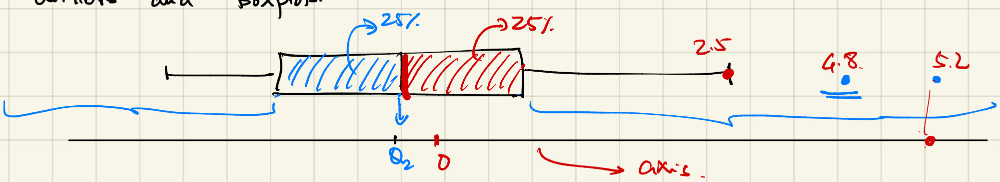
S_n = time until the n^{th} occurence.

Final Exam Review

Chapter 1
- → Sampling
- → Measures of location
  - sample mean
  - sample median
  - percentiles → give quartiles
  - trimmed mean

  } properties:
  ① good for calculations?
  ② Robust to minor changes in data.
- → Measures of scale/spread.
  - → sample variance and std dev.
  - → Interquartile range

Outliers and Boxplots.



Chapter 2 :

- → Axioms of probability
- → calculating probs when all outcomes are equally likely
  - Counting principle
- → Conditional prob. and Bayes.
- → Independence.

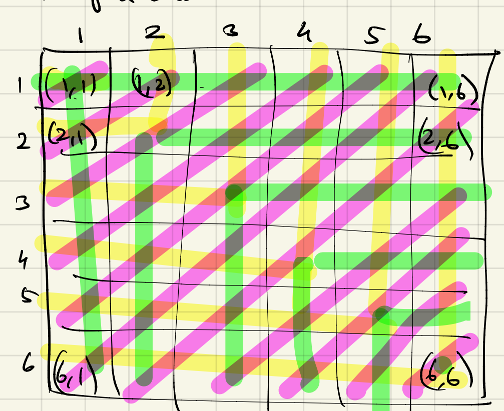
Chapter 3

Random Variable → "procedure to re-describe the sample space S based on certain properties"

X = sum of the dice → 2, 3, 4, 5, 6, ... 12.

Y = max of (x,y)

Rolling a dice two time.



Given X $\longrightarrow$ discrete $\rightsquigarrow$ pmf
$\quad\quad\quad\quad\;\; \hookrightarrow$ Continuous $\rightsquigarrow$ pdf

Calculate $E(X)$, $V(X)$, $\sigma_X = \sqrt{V(X)}$

Discrete R.V's

① Uniform discrete

② Binomial dist (special case is Bernoulli dist)

③ Hypergeo.

④ Piosson dist

⑤ Negative binomial (Geometric dist).

Continuous R.Vs:

① Uniform cont r.v. $[A, B]$

② Normal Dist

③ Gamma Dist $\longrightarrow$ Exp
$\quad\quad\quad\quad\quad\quad \hookrightarrow$ chi sq.

④ Beta, log normal, Cauchy, Weibull.

If $X \sim Bin(n=10, p=0.8)$

$\{X=5\}$ = set of all samples of size 10 that have exactly 5 success.

$\{X \leq 5\}$

$\{X \geq 5\}$

$\{X \neq 5\}$

Chapter 5) Joint Distributions

$$p(x,y) = P(X=x, Y=y)$$

$$\boxed{p(x,y) = p_{X|Y}(x|y) \cdot p_Y(y) = p_{Y|X}(y|x) \cdot p_X(x)}$$

X, Y to independent if

$$p_{X|Y}(x|y) = p_X(x) \qquad \forall x,y \in \mathcal{X} \times \mathcal{Y}$$

Random Samples:

$\{X_1, X_2, \ldots, X_n\} \rightsquigarrow$ X_i's are ind

X_i's have the same dist

iid, independent and identically dist.

Statistic $\rightsquigarrow$

$\hat{\theta}(X_1, \ldots, X_n) =$

Central Limit Theorem. $\rightsquigarrow$ sampling distribution of $\bar{X}$ and T_0.

Gamma: If $X_1, \ldots, X_n$
$\Gamma(\alpha_1, \beta) \quad \Gamma(\alpha_n, \beta)$

Then $\Gamma(\underline{\alpha_1 + \alpha_2 + \cdots + \alpha_n}, \beta)$

Covariance

$$\mathrm{Cov}(X,Y) = E(\underbrace{(X-\mu_X)(Y-\mu_Y)}_{E(h(X,Y))})$$

$$= \sum_x \sum_y (x-\mu_X)(y-\mu_Y) \cdot p(x,y)$$

Linear Combination of R.Vs.

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n.$$

$E(Y)$

$\mathrm{Cov}(X_i, X_i) = V(X_i)$

$$\boxed{V(Y) = \sum_i \sum_j a_i a_j \,\mathrm{Cov}(X_i, X_j)}$$

If $X_1, \ldots, X_n$ are ind.

$$V(Y) = a_1^2 V(X_1) + \cdots + a_n^2 V(X_n)$$

$Y = a_1 X_1 + a_2 X_2$

$$V(Y) = \sum_{i=1}^{2} \sum_{j=1}^{2} a_i a_j \,\mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{2} (a_i a_1 \mathrm{Cov}(X_i, X_1) + a_i a_2 \mathrm{Cov}(X_i, X_2))$$

$a_{ij} = a_i \cdot a_j$

$$= \sum_{i=1}^{2} (a_{i1} \mathrm{Cov}(X_i, X_1) + \sum_{i=1}^{2} a_{i2} \mathrm{Cov}(X_i, X_2)$$

$$= (a_{11} \mathrm{Cov}(X_1, X_1) + a_{12} \mathrm{Cov}(\cancel{X_2, X_1})) + (a_{12} \mathrm{Cov}(\cancel{X_1, X_2}) + a_{22} \mathrm{Cov}(X_2, X_2))$$

$f(x)$, $f(y)$

$X + Y \longrightarrow \dfrac{f(x) + f_Y}{f(x) \cdot f(y)}$?

Chapter 6) Estimators

statistic depending on a r.sample

$Bias(\hat{\theta}) = E(\hat{\theta} - \theta)$

$= E(\hat{\theta}) - \theta.$

Unbiased estimators $\rightarrow$ $Bias(\hat{\theta}) = 0$ $\forall$ possible choices of θ

$V(\hat{\theta})$ = variance of $\hat{\theta}$.

MVUE $\rightarrow$

$MSE(\hat{\theta}) \overset{?}{=} Var(\hat{\theta}) + \underline{Bias(\hat{\theta})^2}$ $\rightarrow$ Bias Variance tradeoff.

$E((\hat{\theta} - \theta)^2)$.

75 percentile.

$P(X \leq \underline{x(p)}) = 0.75$

qnorm(0.75, mean = 30, sd = 7.8m)

$0.67 \rightarrow x = \sigma \times z + \mu$

$= \underline{7.8 \times 0.67 + 30}$

Methods of Calculating Estimators

① Method of Moments

② Maximum likelihood estimates