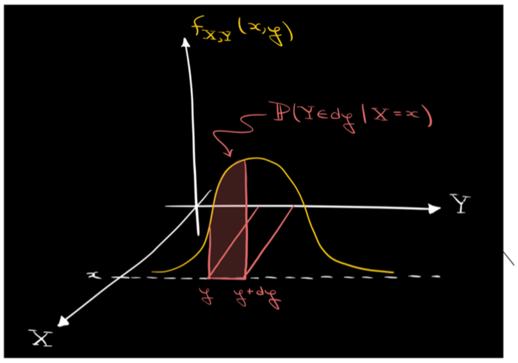
Stat 134 lec 36

Wednesday Quiz 7 Sec 5.4, 6.1, 6.2, 6.3

Last time

sec 6.3 Conditional density



$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

Bayesian statistics

We wish to learn about ~~the~~ distribution of the parameter Θ of Pois(Θ).

We assume $\Theta \sim$ Gamma(r,λ) (Prior) and collect data $N=(N_1, N_2, \dots, N_m)$ where $N_1, \dots, N_m \overset{iid}{\sim}$ Pois(Θ)

Posterior $\propto$ likelihood $\cdot$ Prior

$$P(N_1=n_1, \dots, N_m=n_m \mid \Theta=\theta)$$
$$\parallel$$
$$P(N_1=n_1 \mid \Theta=\theta) \cdots P(N_m=n_m \mid \Theta=\theta)$$

conditional density

$$f_{\Theta|N=(n_1,\dots,n_m)}(\theta) \propto \left(\frac{e^{-\theta}\theta^{n_1}}{n_1!} \cdots \frac{e^{-\theta}\theta^{n_m}}{n_m!}\right) \cdot \theta^{r-1} e^{-\theta\lambda}$$

$$\propto \theta^{n_1+\dots+n_m+r-1} e^{-(m+\lambda)\theta}$$

so $\Theta \mid N=(n_1,\dots,n_m) \sim$ Gamma$(\Sigma n_i + r, m+\lambda)$ ← updated distribution of Θ given data.

Today

sec 6.4 ① covariance and the variance of sum
② correlation

Tentative schedule

M 6.4

W 6.4 / 6.5

F 6.5 HW 13 due, pass out S18 final
HW 14 not to turn in.

RRR week

M 6.5, start review.

W review

F review S18 final

Final exam Th Dec 13 Room TBA.

① Sec 6.4 Covariance and variance of a sum

$X, Y, \quad S = X + Y$

mean $\mu_X, \mu_Y, \mu_S = \mu_X + \mu_Y$

$$\mathrm{Var}(S) = E\left((S-\mu_S)^2\right) = E(D_S^2)$$

$S-\mu_S = D_S$ deviation from mean

$$D_S = S - \mu_S$$
$$= X + Y - (\mu_X + \mu_Y)$$
$$= D_X + D_Y$$

$$\mathrm{Var}(S) = E\left((D_X + D_Y)^2\right)$$
$$= E\left(D_X^2 + D_Y^2 + 2D_X D_Y\right)$$
$$= \underbrace{E(D_X^2)}_{\mathrm{Var}(X)} + \underbrace{E(D_Y^2)}_{\mathrm{Var}(Y)} + 2\underbrace{E(D_X D_Y)}_{\mathrm{Cov}(X,Y)}$$

Def^n The Covariance of X and Y is

$$\mathrm{Cov}(X,Y) = E\left((X-\mu_X)(Y-\mu_Y)\right)$$

Bilinearity Properties

ⓐ $\text{Cov}(X+Y, Z) = \text{Cov}(X,Z) + \text{Cov}(Y,Z)$

ⓑ $\text{Cov}(aX, bY) = ab\,\text{Cov}(X,Y)$

a)

$$\text{Cov}(X+Y, Z) = E\left((X+Y-\mu_{X+Y})(Z-\mu_Z)\right)$$

$$= E\left(\left[(X-\mu_X)+(Y-\mu_Y)\right](Z-\mu_Z)\right)$$

$$= E\left((X-\mu_X)(Z-\mu_Z) + (Y-\mu_Y)(Z-\mu_Z)\right)$$

$$= E\left((X-\mu_X)(Z-\mu_Z)\right) + E\left((Y-\mu_Y)(Z-\mu_Z)\right)$$

$$= \text{Cov}(X,Z) + \text{Cov}(Y,Z).$$

b) $$\text{Cov}(aX, bY) = E\left((aX-\mu_{aX})(bY-\mu_{bY})\right)$$

$$= E\left((aX-a\mu_X)(bY-b\mu_Y)\right)$$

$$= E\left(ab(X-\mu_X)(Y-\mu_Y)\right)$$

$$= ab\,E\left((X-\mu_X)(Y-\mu_Y)\right)$$

$$= ab\,\text{Cov}(X,Y)$$

More generally

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

Proved end of lecture,

Thm $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

Easy facts

$\mathrm{Cov}(X,X) = E(X^2) - E(X)^2 = \mathrm{Var}(X)$

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

$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$

$\mathrm{Cov}(X,c) = 0$

Constant

ex

Simplify

$$\mathrm{Cov}(X-5Y,\ 3X+Y-Z+10)$$

soln

$$3\,\mathrm{Var}(X) + \mathrm{Cov}(X,Y) - \mathrm{Cov}(X,Z) + 0$$

$$-15\,\mathrm{Cov}(X,Y) - 5\,\mathrm{Var}(Y) + 5\,\mathrm{Cov}(Y,Z) + 0$$

Recall X, Y independent

$\Rightarrow \quad E(XY) = E(X)E(Y)$

$\mathrm{Cov}(X, Y) = 0$ if X, Y independent.

Hence if X, Y indep,

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\underbrace{\mathrm{Cov}(X,Y)}_{=0}$$

$$= \boxed{\mathrm{Var}(X) + \mathrm{Var}(Y)}$$

ex

Let $X_1, \dots, X_n$ be identically distributed

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

variance-covariance matrix has all n^2 terms

$$\begin{array}{c|c} & \begin{matrix} X_1 & \quad X_2 & \quad \cdots & \quad X_n \end{matrix} \\ \begin{matrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{matrix} & \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) & & \\ & \mathrm{Cov}(X_2, X_2) & & \\ & & \ddots & \\ & & & \mathrm{Cov}(X_n, X_n) \end{bmatrix}_{n \times n} \end{array}$$

$$\mathrm{var}\left(\sum_{i=1}^{n} X_i\right) = n\,\mathrm{var}(X_1) + n(n-1)\,\mathrm{Cov}(X_1, X_2)$$

diagonal — off diagonal

ex

$x_1, \dots, x_N$ — pop mean μ, pop SD σ

drawn n w/o replacement (called a simple random sample (SRS))

$X_1, \dots, X_n$

let $S_n = X_1 + \dots + X_n$ be the sample mean

Find $E(S_n)$

$\mathrm{Var}(S_n)$.

$$E(S_n) = E(X_1 + \dots + X_n) = nE(X_1) = n\mu$$

$$\mathrm{Var}(S_n) = n\underbrace{\mathrm{Var}(X_1)}_{=\sigma^2} + n(n-1)\,\mathrm{Cov}(X_1, X_2)$$

trick to find $\mathrm{Cov}(X_1, X_2)$

take a census $n = N$

$$\mathrm{var}(S_n) = \mathrm{var}(S_N) = 0$$

$$\Rightarrow 0 = N\sigma^2 + N(N-1)\,\mathrm{Cov}(X_1, X_2)$$

$$\Rightarrow \operatorname{Cov}(X_1, X_2) = -\frac{N\sigma^2}{N(N-1)} = \boxed{-\frac{\sigma^2}{N-1}}$$

So
$$\operatorname{Var}(S_n) = n\sigma^2 + n(n-1)\left(-\frac{\sigma^2}{N-1}\right)$$
$$= n\sigma^2\left[\frac{N-n}{N-1}\right]$$

correction factor < 1

(note if draw w/ replacement
$\operatorname{Var}(S_n) = n\sigma^2$).

Recall, we have seen this correction factor before (p241)

$$X \sim \operatorname{Bin}(n, p) \Rightarrow \operatorname{Var}(X) = (npq)$$

$$X \sim \text{hyper-geom}(n, N, G) \Rightarrow \operatorname{Var}(X) = n\left(\frac{G}{N}\right)\left(\frac{N-G}{N}\right)\left(\frac{N-n}{N-1}\right).$$

p q correction factor

Sec 6.4 Correlation

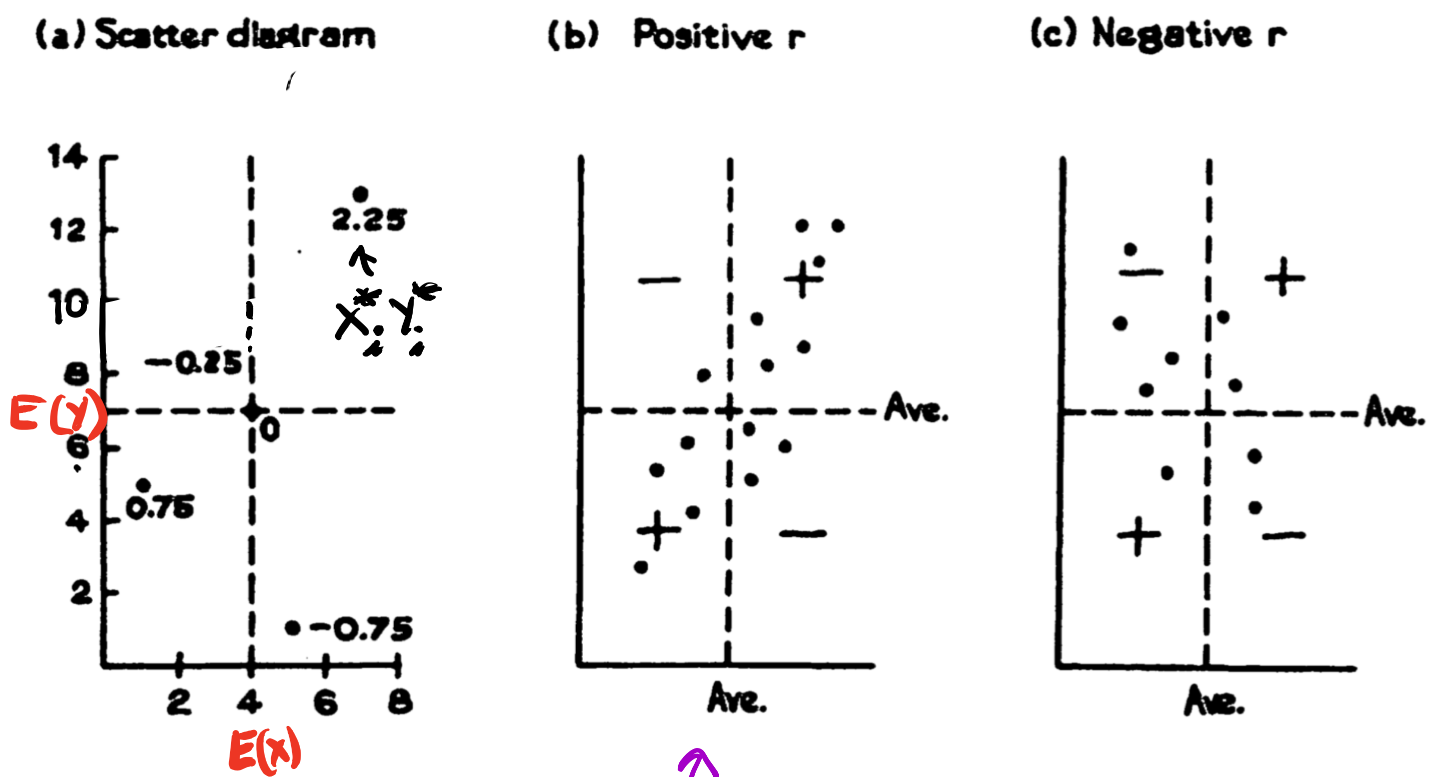
$$Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$$

$$r = Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)} = E\left(\left(\frac{X-\mu_X}{SD_X}\right)\left(\frac{Y-\mu_Y}{SD_Y}\right)\right)$$

$$= E(X^* Y^*)$$

X, Y in standard units

How the correlation coefficient works.



This will have a positive correlation since more of the points are in the 1st and 3rd quadrants ex Suppose the sum of k exchangeable (i.e identically distributed) RVs is a constant

$$N_1 + N_2 + \cdots + N_k = c$$

Find $\mathrm{Corr}(N_1, N_2)$.

soln

$$N_1 + N_2 + \cdots + N_k = c$$

$$\Rightarrow \mathrm{Var}(N_1 + \cdots + N_k) = 0$$

$$\Rightarrow k\,\mathrm{Var}(N_1) + k(k-1)\,\mathrm{Cov}(N_1, N_2) = 0$$

$$\Rightarrow \mathrm{Cov}(N_1, N_2) = \frac{-\mathrm{Var}(N_1)}{k-1}$$

$$\Rightarrow \mathrm{Corr}(N_1, N_2) = \frac{\mathrm{Cov}(N_1, N_2)}{\underbrace{SD(N_1)\,SD(N_2)}_{= \mathrm{Var}(N_1)}} = \boxed{-\frac{1}{k-1}}$$

concept test

Stat 134

Monday Nov 26 2018

1. Consider a Poisson(λ) process. Let $T_r \sim$ gamma(r, λ) be the rth arrival time. Cov(T_1, T_3) equals:

   a λ

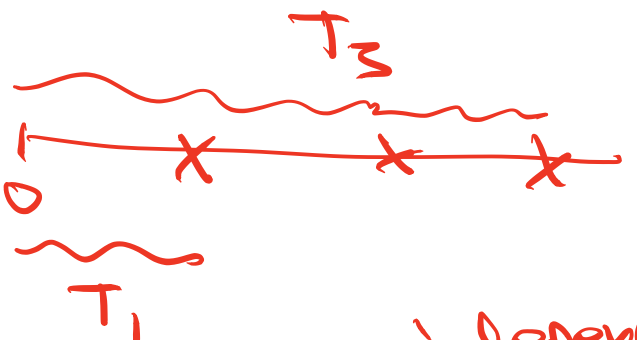
   b λ^2

   c $1/\lambda^2$

   d none of the above

Recall $\quad Var(T_r) = \frac{r}{\lambda^2}$

soln



independence

$$Cov(T_1, T_3 - T_1) = 0$$

$$Cov(T_1, T_3) - Var(T_1) = 0$$

$$Cov(T_1, T_3) = Var(T_1) = \boxed{\frac{1}{\lambda^2}}$$

2. Consider a Poisson(λ) process. Let $T_r \sim$ gamma(r, λ) be the rth arrival time. Corr(T_1, T_3) equals:

a $\sqrt{\frac{3}{1}}$

b $\sqrt{\frac{1}{3}}$

c $\sqrt{\frac{2}{3}}$

d none of the above

$$\mathrm{Corr}(T_1, T_3) = \frac{\mathrm{Cov}(T_1, T_3)}{SD(T_1)\, SD(T_3)} = \frac{\frac{1}{\lambda^2}}{\frac{1}{\lambda} \cdot \frac{\sqrt{3}}{\lambda}} = \frac{1}{\sqrt{3}}$$

Appendix

Thm $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

pf/ $$\mathrm{Cov}(X,Y) = E(D_X D_Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$
$$= E(XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y)$$
$$= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y)$$
$$= E(XY) - E[X]E(Y)$$

$\square$