Week 7

- matching problem
- best prize "
- more on the exponential
- ageless/memoryless
- failure/hazard rate functions

eg <u>Matching problem</u>

n envelopes

n letters

letters to envelopes at random

$$P(\underbrace{j \text{ matches}}_{A_{nj}})?$$

<u>Proposition</u> $$P(A_{nj}) = \frac{1}{j!} \sum_{k=j}^{n} \frac{(-1)^{k-j}}{(k-j)!}$$

<u>Proof</u>: Let $I_i = I(\{i\text{th letter matches}\})$
$\sim \text{Bernoulli}(\frac{1}{n})$

Let P_j = prob that a fixed set of j letters matches & the others don't

Then $$P(A_{nj}) = \binom{n}{j} P_j$$

Only look at first j.

$$Q_j = P(\text{1st } j \text{ match \& others do what they want})$$

$$= \frac{(n-j)!}{n!}$$

$$P_j = E\left[\prod_{i=1}^{j} I_i \prod_{h=j+1}^{n} (1-I_h)\right]$$

&

$$Q_j = E\left(\prod_{i=1}^{j} I_i\right)$$

Now $\displaystyle\prod_{h=j+1}^{n} (1-I_h) = \sum_{m=0}^{n-j} (-1)^m S_m$,

where S_m = sum of $\binom{n-j}{m}$ products of the I_h's $(h=j+1,\dots,n)$ & $S_0 = 1$. Hence

$$\prod_{i=1}^{j} I_i \prod_{h=j+1}^{n} (1-I_h) = \overbrace{\sum_{m=0}^{n-j} (-1)^m S_{m+j}}$$

$$\Rightarrow P_j = \sum_{m=0}^{n-j} (-1)^m \overbrace{Q_{m+j}\binom{n-j}{m}}^{= E(S_{m+j})}$$

$$= \sum_{m=0}^{n-j} (-1)^m \binom{n-j}{m} \frac{(n-(j+m))!}{n!}$$

$$\overset{k=m+j}{=} \sum_{k=j}^{n} (-1)^{k-j} \frac{(n-j)!}{(k-j)!(n-k)!} \frac{(n-k)!}{n!}$$

$$= \frac{(n-j)!}{n!} \sum_{k=j}^{n} \frac{(-1)^{k-j}}{(k-j)!}$$

$$\Rightarrow P(A_{n_j}) = \binom{n}{j} \times \uparrow = \frac{1}{j!} \sum_{k=j}^{n} \frac{(-1)^{k-j}}{(k-j)!}$$

qed

<u>Note</u> $P(A_{n0}) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \approx e^{-1}$

$= P(\text{Poisson}(1) = 0)$

Not surprising as $X = \sum_{i=1}^{n} I_i$

$\uparrow$ Bernoulli$(\frac{1}{n})$ but dependent

<u>Note</u> $E(I_i I_j) = P(i\text{th} \,\&\, j\text{th match})$

$i \neq j$

$= \underbrace{P(j\text{th matches} \mid i\text{th matches})}_{\frac{1}{n-1}} \underbrace{P(i\text{th matches})}_{\frac{1}{n}}$

$\Rightarrow \text{cov}(\quad) = \frac{1}{n(n-1)} - \frac{1}{n^2} \rightarrow \text{Var}(X)$

Alternative sol'n to the matching problem

Let X_n = # of matches (in the "n case")

Set
$$G_n(s) = E(s^{X_n})$$
$$= \sum_{j=0}^{n} s^j P(X_n = j)$$
$$= \sum_{j=0}^{n} s^j P(A_{nj})$$

Let M_{nj} = # of arrangements leading to j matches
$$= n!\, P(A_{nj})$$

Now, given j matches the probability that the last letter (the nth one) is a match is j/n. Hence

$\frac{j}{n} M_{nj}$ = # of arrangements leading to j matches with the nth letter matched

Clearly
$$\frac{j+1}{n+1} M_{n+1,\, j+1} = M_{nj}$$

$$\Rightarrow \frac{M_{mj}}{m!} = \frac{j+1}{m+1} \frac{M_{m+1,j+1}}{m!} = (j+1)\frac{M_{m+1,j+1}}{(m+1)!}$$

$$\Rightarrow P(A_{mj}) = (j+1) P(A_{m+1,j+1})$$

$$\Rightarrow \sum_{j=0}^{m} s^j P(A_{mj}) = \sum_{j=0}^{m} s^j (j+1) P(A_{m+1,j+1})$$

$$= \sum_{j=0}^{m} \left(\frac{d}{ds} s^{j+1}\right) P(A_{m+1,j+1})$$

$$= \frac{d}{ds} \sum_{j=0}^{m} s^{j+1} P(A_{m+1,j+1})$$

$$= \frac{d}{ds} \sum_{k=1}^{m+1} s^k P(A_{m+1,k}) = \frac{d}{ds} \sum_{k=0}^{m+1} s^k P(A_{m+1,k})$$

(derivative of $k=0$ term $=0$)

$$= \frac{d}{ds} G_{m+1}(s)$$

Hence

$$G'_{m+1}(s) = G_m(s) \quad , m = 1, \dots$$

+ since $X_1 = 1$ we have $G_1(s) = s$.

So

$$G_{m+1}(s) = \int_1^s G_m(u)\,du + c$$

Since $G_{m+1}(1) = 1$ (true $\forall$ pgf's) we see $c = 1$ & so

$$\left.\begin{aligned} G_{m+1}(s) &= 1 + \int_1^s G_m(u)\,du \quad , m = 1, 2, \cdots \\ G_1(s) &= s \end{aligned}\right\} \quad (**)$$

Now use induction, or otherwise, to conclude that the solution of (**) is

$$G_m(s) = \sum_{k=0}^{m} \frac{(s-1)^k}{k!} \quad , m = 1, 2, \cdots$$

$\Rightarrow$ the result (the coefficient of s^j is as stated)

Notice $\sum_{k=0}^{m} \frac{(s-1)^k}{k!} \to \underbrace{e^{(s-1)}}_{\text{pgf of a Poisson}(1)}$ as $m \to \infty$

eg <u>Best prize problem</u>

- n prizes (ordered from top to bottom)
- there is a best one
- don't know them but n is known
- pick a prize at random

keep / throw away + pick again (that prize is lost)

P(get best prize)? ← depends on how you go along the keep + throwing away

<u>Fix k where $0 \le k \le n$</u>

throw away 1st k + pick the next one which is better than the 1st k.

<u>Sol'n</u> Let X = order/position of the best prize $\in \{1, 2, \cdots, n\}$

$$P_k(\text{best}) = \sum_{i=1}^{n} P_k(\text{best} \mid X=i)\, P(X=i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} P_k(\text{best} \mid X=i)$$

Clearly $P_k(\text{best} \mid X=i) = 0$, if $i \le k$

If $i > k$

$P_k(\text{best} \mid X=i) = P(k+1, \cdots, i-1$ prizes are not as good as the 1st $k)$

$= P($ best of the first $i-1$ is in the first $k)$

$$= \frac{k}{i-1}$$

$$\therefore P_k(\text{best}) = \frac{1}{n} \sum_{i=k+1}^{n} \frac{k}{i-1}$$

$$= \frac{k}{n} \sum_{i=k+1}^{n} \frac{1}{i-1}$$

$$= \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j}$$

$$\simeq \frac{k}{n} \int_{k}^{n-1} \frac{1}{x}\, dx$$

$$= \frac{k}{n} \log\left(\frac{n-1}{k}\right) \simeq \frac{k}{n} \log\left(\frac{n}{k}\right)$$

Look at

$$g(x) = \frac{x}{n} \log\left(\frac{n}{x}\right)$$

Then $g'(x) = \frac{1}{n} \log\left(\frac{n}{x}\right) - \frac{1}{n}$

$$\Rightarrow x = \frac{n}{e}$$

$$\& \ g(x) = \frac{1}{e}$$

So $k \approx \frac{n}{e}$ & then $P(\text{best}) \approx \frac{1}{e} \approx \frac{1}{2.7}$

$> 35\%$

The Exponential & related matters

Look at $X \geq 0$.

Ageless or memoryless property

If $P(X > s+t \mid X > s) = P(X > t)$, $\forall\, s,t \geq 0$

then X is ageless.

$F(x) = P(X \leq x)$

$\bar{F}(x) = P(X > x) = 1 - F(x)$

Either F or $\bar{F}$ determines the distribution (= collection of all probabilities or expectations)

If we assume a pdf $f(x)$

then $$\lambda(x) = \frac{f(x)}{\bar{F}(x)}$$

is the <u>hazard</u> or <u>failure rate</u> function.

$$\lambda(x)dx = \frac{f(x)dx}{\bar{F}(x)}$$

$$= \frac{P(X \in (x, x+dx))}{P(X > x)}$$

$$= P(X \in (x, x+dx) / X > x)$$

<u>Note</u> ① The ageless property is

$$\bar{F}(s+t) = \bar{F}(s)\,\bar{F}(t), \quad \forall s,t \geq 0$$

$$\Longrightarrow \bar{F}(x) = e^{-\lambda x}, \quad x \geq 0$$

for some $\lambda > 0$.

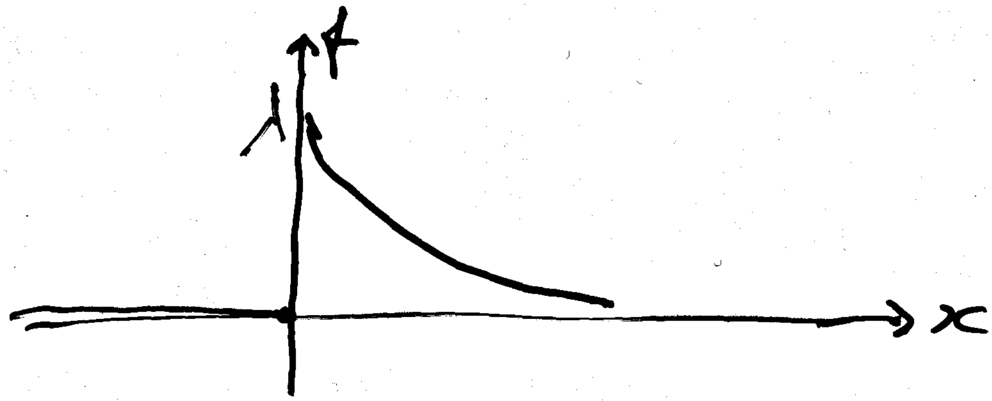
② If X is a cts rv then $F'(x) = f(x)$ & $\bar{F}'(x) = -f(x)$

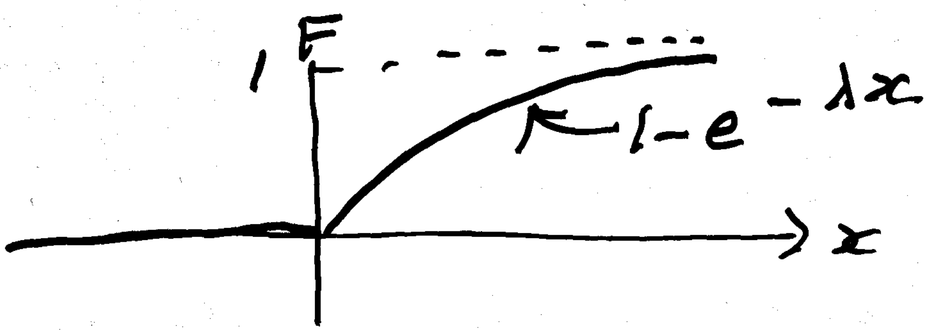
③ ① says that the only ageless

rv has pdf

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$
$$= 0, \quad \text{ow}$$

& this is the pdf of the exponential(λ) distribution.





④ If X is ageless then $\lambda(x) = \lambda$

⑤ $\lambda(x)$ determines $\bar{F}$

$$\lambda(x) = \frac{f(x)}{\bar{F}(x)} = -\frac{\bar{F}'(x)}{\bar{F}(x)} = -\frac{d}{dx}\log[\bar{F}(x)]$$

$$\Longrightarrow \log[\bar{F}(x)] = -\int_0^x \lambda(u)\,du + \text{constant}$$

Now $\bar{F}(0) = P(X>0) = 1$ &

$\log(1) = 0$ & so

$$\log[\bar{F}(x)] = -\int_0^x \lambda(u)\,du$$

$$\Rightarrow \quad \bar{F}(x) = e^{-\int_0^x \lambda(u)\,du}$$

In particular if $\lambda(u) = \lambda$ then

$$\bar{F}(x) = e^{-\lambda x}$$

& so the distribution must be exponential(λ).

<u>Moment generating function</u> (mgf)

The mgf of X is

$$m(t) = E(e^{tX})$$

<u>Remark</u> If $m(t)$ exists in a neighborhood of $t=0 \Rightarrow$ we know the dist'n.

eg Let $X \sim$ exponential (λ).
Its mgf is

$$m(t) = \int_{-\infty}^{\infty} f(x) e^{tx} dx$$

$$= \int_{0}^{\infty} e^{-\lambda x} e^{tx} dx$$

$$= \frac{\lambda}{\lambda - t} \quad , \quad t < \lambda$$

Notice

$$m'(t) = \frac{\lambda}{(\lambda - t)^2}$$

$$\Rightarrow m'(0) = \frac{1}{\lambda} = E(X)$$

$$m''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

$$\Rightarrow E(X^2) = \frac{2}{\lambda^2}$$

$\therefore$ Var(X)

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

If we add r iid exponential(λ) rv's then the mgf of the sum is

$$\left(\frac{\lambda}{\lambda-t}\right)^{r}, \quad t<\lambda$$

Now consider the rv Y with pdf

$$f(y) \propto y^{r-1}e^{-\lambda y}, \quad y>0$$
$$= 0 \quad , \text{ow}$$

$\propto$ means "proportional to". In this case

$$f(x) = c\,y^{r-1}e^{-\lambda y}, \quad y>0$$

c has to satisfy

$$\int_0^\infty c\,y^{r-1}e^{-\lambda y}dy = 1$$

$$\overset{u=\lambda y}{=} \int_0^\infty c\left(\frac{u}{\lambda}\right)^{r-1}e^{-u}\frac{1}{\lambda}\,du$$

$$= \int_0^\infty c \, \frac{u^{r-1} e^{-u}}{\lambda^r} du = 1$$

$$\Rightarrow c = \frac{\lambda^r}{\int_0^\infty u^{r-1} e^{-u} du}$$

$$= \frac{\lambda^r}{(r-1)!}$$

So the pdf that we are looking at is

$$f(y) = \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y} \quad , y > 0$$

$$= 0 \quad , ow$$

The mgf of this is

$$\int_0^\infty \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y} e^{ty} dy$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^r \quad , t < \lambda$$

Hence the pdf of a sum of r iid exponential(λ)'s is

$$\left.\begin{aligned} f(y) &\propto y^{r-1} e^{-\lambda y}, && y>0 \\ &= 0, && \text{ow} \end{aligned}\right\} (*)$$

If $r>0$ then $(*)$ continues to be a pdf if we take the constant to be

$$\frac{\lambda^r}{\Gamma(r)}$$

Note $(*)$ is the pdf of a gamma(r, λ)

where

$$\Gamma(r) = \int_0^\infty u^{r-1} e^{-u}\, du$$

is the gamma function.

Note

$$\left.\begin{aligned} &- \Gamma(1) = 1 \\ &- \Gamma(r+1) = r\,\Gamma(r) \end{aligned}\right\} \Rightarrow \Gamma(n+1) = n!$$

$$- \Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$