Lecture 23: Mar 2, Conditional pmf, pdf, cdf Ross 6.4, 6.5

23.1 Review from Lecture 12

(i) For any two discrete random variables X and W, we considered the conditional probability mass function $p_{X|W}(x|w) \;=\; P(X=x|W=w) \;=\; p_{X,W}(x,w)/p_W(w)$, for w such that $p_W(w)>0$.

(ii) We considered the case $X \sim \mathcal{P}o(\mu)$ and $Y \sim \mathcal{P}o(\nu)$, and X,Y independent, so $W=(X+Y) \sim \mathcal{P}o(\mu+\nu)$. Then we considered $P(X=x \mid W=w) \;=\; P(X=x, Y=w-x)/P(W=w) \;=\; p_X(x)p_Y(w-x)/P(W=w)$ and we showed $X \mid (X+Y)=w$ is Binomial $(w,\ \mu/(\mu+\nu))$.

(iii) We **defined** the conditional pdf $f_{X|W}(x|w) \;=\; f_{X,W}(x,w)/f_W(w)$, for w such that $f_W(w)>0$. This definition is motivated by

$$P(x<X\le x+\delta x \mid w<W\le w+\delta w) \;=\; \frac{P(x<X\le x+\delta x \cap w<W\le +\delta w)}{P(w<W\le w+\delta w)} \approx \frac{f_{X,W}(x,w)\ \delta x\ \delta w}{f_W(w)\ \delta w}$$

(iv) We considered the case $X \sim G(\alpha_1,\lambda)$, $W \sim G(\alpha_2,\lambda)$, X and Y independent, so $W\equiv X+Y \sim G(\alpha_1+\alpha_2,\lambda)$. (Actually, then we could only consider integer $\alpha_1=m$,$\alpha_2=n$, but now we can have any $\alpha_1,\alpha_2>0$.) Then we considered $f_{X|W}(x|w) \;=\; f_{X,W}(x,w)/f_W(w) \;=\; f_X(x)\ f_Y(w-x)/f_W(w)$ and we showed $X \mid (X+Y)=w$ has pdf

$$f_{X|(X+Y)}(x|w) \;=\; \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}w^{-1}(x/w)^{m-1}(1-(x/w))^{n-1} \quad \text{on } 0\le x\le w \quad \text{and 0 otherwise}$$

23.2 The conditional pmf/pdf is a pdf/pmf.

(i) Recall back in 394 (Ross 3.5) we showed a conditional probability **is** a probability; obeys all the axioms and resulting formulae. Now we do the same for the conditional pmf/pdf.

pmf: $p_{X|W}(x|w)\ge 0$, and $\sum_{x\in\mathcal{X}} p_{X|W}(x|w) \;=\; \sum_x p_{X,W}(x,w)/p_W(w) \;=\; p_W(w)/p_W(w) \;=\; 1$.

pdf: $f_{X|W}(x|w)\ge 0$ and $\int_x f_{X|W}(x|w)dx \;=\; \int_x f_{X,W}(x,w)\ dx/f_W(w) \;=\; f_W(w)/f_W(w) \;=\; 1$.

(ii) Hence there is a conditional cdf (Ross, P.292). By definition, $P(X\in A \mid w) \;=\; \int_A f_{X|W}(x|w)\ dx$, so with $A=(-\infty,a]$, $F_{X|W}(a|w) \;=\; P(X \;\le\; a|W=w) \;=\; \int_{-\infty}^{a} f_{X|W}(x|w)\ dx$.

(iii) From joint to and from conditional pmf/pmf

For events A,B, $P(A|B)\ P(B) \;=\; P(A\cap B) \;=\; P(B|A)\ P(A)$.

For discrete random variables: $p_{X|Y}(x|y)p_Y(y) \;=\; p_{X,Y}(x,y) \;=\; p_{Y|X}(y|x)p_X(x)$.

For continuous random variables: $f_{X|Y}(x|y)f_Y(y) \;=\; f_{X,Y}(x,y) \;=\; f_{Y|X}(y|x)f_X(x)$.

23.3 Independence via conditional pmf/pdf

For events A,B, $P(A\cap B) \;=\; P(A).P(B)$; $P(B|A)=P(B)$ and $P(A|B)=P(A)$.

For discrete random variables: $p_{X,Y}(x,y) \;=\; p_X(x)p_Y(y)$; $p_{X|Y}(x|y)=p_X(x)$ and $p_{Y|X}(y|x)=p_Y(y)$.

For continuous random variables: $f_{X,Y}(x,y)=f_X(x)f_Y(y)$; $f_{X|Y}(x|y)=f_X(x)$ and $f_{Y|X}(y|x)=f_Y(y)$.

If we find $f_{X|Y}(x|y)$ does not depend on y, they X and Y are independent.

23.4 Example of the independent Gammas and their sum

$$f_{X|W)}(x|w) \;=\; \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}w^{-1}(x/w)^{m-1}(1-(x/w))^{n-1} \quad \text{on } 0\le x\le w \quad \text{and 0 otherwise}$$

Recall for a scale parameter: $X=wV$; $f_X(x)=w^{-1}f_W(x/w)$, so here $V=X/w$ given $W=w$ has pdf $f_{V|W)}(x|w) \;=\; (\Gamma(m+n)/\Gamma(m)\Gamma(n))v^{m-1}(1-v)^{n-1}$ on $0\le v\le 1$ and 0 otherwise.

But this does not depend on w!! i.e. we have shown $V=X/(X+Y)$ is independent of $W=(X+Y)$.

Lecture 24: Mar 4, Conditional Expectations, and their Expectations Ross 7.5

24.1 Conditional expectation as a random variable

(i) A conditional pdf gives rise to conditional expectations:

$$\mathrm{E}(X \mid Y = y) \;=\; \int_{x=-\infty}^{\infty} x \; f_{X|Y}(x|Y = y) \; dy$$

Similarly for a discrete X with conditional pmf, with sums replacing integrals.

(ii) Conditional expectations satisfy usual properties:

$$\mathrm{E}(g(X) \mid Y = y) \;=\; \int_{x=-\infty}^{\infty} g(x) \; f_{X|Y}(x|Y = y) \; dy$$

$$\mathrm{E}(g(X,Y) \mid Y = y) \;=\; \int_{x=-\infty}^{\infty} g(x,y) f_{X|Y}(x|Y = y) \; dy$$

assuming integrals/sums converge absolutely.

(iii) Conditional expectations satisfy usual properties: $\mathrm{E}(\sum_i X_i \mid Y = y) \;=\; \sum_i \mathrm{E}(X_i \mid Y = y)$.

24.2 The expectation of a conditional expectation

The expressions 24.1 depend on the value y; that is they are functions of y. We can consider the corresponding function of Y. That is $\mathrm{E}(X|Y)$ is a random variable; a function of Y. Then

$$\begin{aligned}\mathrm{E}(\mathrm{E}(X|Y)) &= \int_y \mathrm{E}(X|Y = y) \; f_Y(y) \; dy \;=\; \int_y \left(\int_x x \; f_{X|Y}(x|y) \; dx \right) f_Y(y) \; dy \\ &= \int_x \int_y x \; f_{X,Y}(x,y) \; dy \; dx \;=\; \int_x x \; f_X(x) \; dx \;=\; \mathrm{E}(X).\end{aligned}$$

Similarly: $\mathrm{E}(\mathrm{E}(g(X) \mid Y)) \;=\; \mathrm{E}(g(X))$.

24.3 Conditional variance

(i) $\mathrm{var}(X \mid Y = y) \;=\; \mathrm{E}((X - \mathrm{E}(X|Y = y))^2 \mid Y = y)$ so we define $\mathrm{var}(X \mid Y) \;=\; \mathrm{E}((X - \mathrm{E}(X|Y))^2 \mid Y)$.

(ii) $\mathrm{var}(X \mid Y) \;=\; \mathrm{E}(X^2 \mid Y) - (\mathrm{E}(X|Y))^2$, by multiplying out the expression in (i).

(Note $\mathrm{E}(g_1(X)g_2(Y) \mid Y) \;=\; g_2(Y)\mathrm{E}(g_1(X) \mid Y)$.)

(iii) $\mathrm{E}(\mathrm{var}(X \mid Y)) \;=\; \mathrm{E}(\mathrm{E}(X^2|Y)) \;-\; \mathrm{E}((\mathrm{E}(X|Y))^2) \;=\; \mathrm{E}(X^2) \;-\; \mathrm{E}((\mathrm{E}(X|Y))^2)$

and $\mathrm{var}(\mathrm{E}(X \mid Y)) \;=\; \mathrm{E}((\mathrm{E}(X|Y))^2) - (\mathrm{E}(\mathrm{E}(X|Y)))^2 \;=\; \mathrm{E}((\mathrm{E}(X|Y))^2) - (\mathrm{E}(X))^2$

So $\mathrm{var}(X) \;=\; \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \;=\; \mathrm{E}(\mathrm{var}(X|Y)) \;+\; \mathrm{var}(\mathrm{E}(X|Y))$.

24.4 Example (Ross, P.381)

This example combined a pdf and a pmf: X is discrete, but Y continuous.

Suppose the train arrives to pick up passengers at a time Y Uniformly distributed between 8:00am and τ minutes after 8:00a.m.: $\mathrm{E}(Y) = \tau/2$, $\mathrm{var}(Y) = \tau^2/12$ (in minutes after 8:00a.m.).

Now suppose passengers arrive to catch the train as a Poisson process rate λ, starting at time 7:45 a.m. (when the previous train left). So the number arriving by time y is Poisson mean $\lambda(15 + y)$;

$\mathrm{E}(X \mid Y = y) \;=\; \mathrm{var}(X \mid Y = y) \;=\; \lambda(15 + y)$.

Find the mean and variance of the number of passengers, X, who get on the train:

$\mathrm{E}(X) \;=\; \mathrm{E}(\mathrm{E}(X|Y)) \;=\; \mathrm{E}(\lambda(15 + Y)) \;=\; \lambda(15 \;+\; \tau/2)$.

$\mathrm{var}(X) \;=\; \mathrm{E}(\mathrm{var}(X|Y)) \;+\; \mathrm{var}(\mathrm{E}(X|Y)) \;=\; \mathrm{E}(\lambda(15 + Y)) \;+\; \mathrm{var}(\lambda(15 + Y)) \;=\; \lambda(15 \;+\; \tau/2) \;+\; \lambda^2\tau^2/12$

Lecture 25: Mar 6, Examples of conditional distributions and expectations

25.1 Ross Ch. 6 # 42 $f_{X,Y}(x,y) \;=\; x\;\exp(-x(y+1))$ on $0<x<\infty$, $0<y<\infty$.
Now $f_{X,Y}(x,y) \;\equiv\; f_X(x) f_{Y|X}(y|x)$ so $f_{Y|X}(y|x) \;\propto\; \exp(-xy)$ as function of y.
i.e $f_{Y|X}(y|x) \;=\; x\exp(-xy)$ on $0<y<\infty$: $(Y|X=x) \;\sim\; \mathcal{E}(x)$: $\mathrm{E}(Y|X) \;=\; 1/X$.
Then also $f_X(x) \;=\; \exp(-x)$ on $0<x<\infty$; $X \;\sim\; \mathcal{E}(1)$.
Note $\mathrm{E}(XY) \;=\; \mathrm{E}(\mathrm{E}(XY \mid X)) \;=\; \mathrm{E}(X\mathrm{E}(Y \mid X)) \;=\; \mathrm{E}(X \times 1/X) \;\equiv\; 1$.
Also $(xY \mid X=x) \;\sim\; \mathcal{E}(x/x) \;\equiv \mathcal{E}(1)$ which does not depend on x. i.e. $Z = XY \;\sim \mathcal{E}(1)$ independent of X.

25.2 The bivariate Normal density

Recall this is:

$$f_{X,Y}(x,y) \;=\; (2\pi\sigma\tau\sqrt{1-\rho^2})^{-1}\exp\left(-\frac{1}{2(1-\rho^2)}\left((\frac{x-\mu}{\sigma})^2 \right.\right.$$
$$\left.\left. -2\rho\frac{(x-\mu)(y-\nu)}{\sigma\tau} \;+\; (\frac{y-\nu}{\tau})^2\right)\right)$$

So to find $f_{Y|x}(y|x)$ we just need to extract the terms that depend on y:

$$\begin{aligned} f_{Y|X}(y|x) &\propto \exp\left(-\frac{1}{2(1-\rho^2)}\left((\frac{y-\nu}{\tau})^2 - 2\rho\frac{(x-\mu)(y-\nu)}{\sigma\tau}\right)\right) \\ &\propto \exp\left(-\frac{1}{2(1-\rho^2)\tau^2}(y^2 - 2y(\nu+\rho(\tau/\sigma)(x-\mu)))\right) \end{aligned}$$

So $(Y|X=x)$ is Normal with mean $(\nu+\rho(\tau/\sigma)(x-\mu))$ and variance $(1-\rho^2)\tau^2$.

25.3 Multinomial: Ross P.290, 373

In fact, conditioned multinomial is multinomial (Ross P.290); that is, if $Y_1,...,Y_K$ is $Mn(n,(p_1,...,p_K))$ then given $Y_{k+1}+...+Y_K \;=\; m$, $(Y_1,...,Y_k)$ is $Mn(n-m,(p_1^*,...,p_k^*))$ where $p_j^* = p_j/(p_1+...+p_k)$, for $j=1,...,k$.
Find $\mathrm{E}(Y_j \mid Y_i > 0)$: $\qquad \mathrm{E}(Y_j) \;=\; \mathrm{E}(Y_j \mid Y_i = 0)P(Y_i=0) \;+\; \mathrm{E}(Y_j \mid Y_i > 0)P(Y_i > 0)$.
Now $\mathrm{E}(Y_j) = np_j$, and $\mathrm{E}(Y_j \mid Y_i = 0) \;=\; np_j/(1-p_i)$ (why?). Also $1 - P(Y_i > 0) \;=\; P(Y_i = 0) \;=\; (1-p_i)^n$.
So $np_j \;=\; (np_j/(1-p_i))(1-p_i)^n \;+\; \mathrm{E}(Y_j \mid Y_i = 0)(1-(1-p_i)^n)$
or $\mathrm{E}(Y_j \mid Y_i > 0) \;=\; np_j(1-(1-p_i)^{n-1})/(1-(1-p_i)^n)$.

25.4 (Almost) Midterm example done a different way!!

The jointly continuous random variables X and Y have joint density function

$$f_{X,Y}(x,y) = 2 \text{ on } 0 \le x \le y \le 1 \text{ and } f_{X,Y}(x,y) = 0 \text{ for all other } (x,y).$$

The marginal pdf $f_Y(y)$ of Y is $f_Y(y) = 2y$ on $0<y<1$.
The conditional pdf of $X|Y$ is $f_{X|Y}(x|Y=y) = 1/y$ on $0 \le x \le y$.
i.e. Given $Y=y$, X is Uniform on $(0,y)$.
(i) So what is $\mathrm{E}(X|Y)$?
(ii) Show that $\mathrm{E}(XY) = (1/2)\mathrm{E}(Y^2)$ and hence $\mathrm{cov}(X,Y) = (1/2)\mathrm{var}(Y)$.
(iii) Note by symmetry $\mathrm{var}(X) = \mathrm{var}(1-Y) \;=\; \mathrm{var}(Y)$.
(iv) What is the correlation between X and Y ?