Stat 2911 Lecture Notes

Class 24, 2017

Uri Keich

Conditional density, Sampling from 2-d distributions, Sum of jointly continuous RVs, Sum of independent Gamma RVs, Sum of independent exponential RVs (Rice 3.5, 3.6)

Conditional Dist.

Recall: if $P(B)>0$ then $P(A|B)=\dfrac{P(A\cap B)}{P(B)}$.

If X and Y are discrete RVs with joint pmf p_{XY}, then for $y\in\mathbb{R}$ with $p_Y(y)>0$ we defined

$$p_{X|Y}(x|y)=\frac{p_{XY}(x,y)}{p_Y(y)}.$$

Suppose X and Y have a joint density f_{XY}.

We want to make sense of $F_{X|Y}(x|y)=P(X\le x\,|\,Y=y)$.

However, $P(Y=y)=0$ so we need to be creative.

Assume that f_{XY} is a cont. function and therefore that f_Y is also a cont. fn.

Assume further that $f_Y(y)>0$ and as it is cont. for $\delta>0$,

$$P(Y\in[y,y+\delta])=\int_y^{y+\delta} f_Y(t)\,dt>0.$$

Let I be a finite close interval.

$$\Rightarrow P(X\in I\,|\,y\le Y\le y+\delta)=\frac{P(X\in I,\,Y\in[y,y+\delta])}{P(Y\in[y,y+\delta])}$$

is well defined

$$=\frac{\int_I\int_y^{y+\delta} f_{XY}(x,t)\,dt\,dx}{\int_y^{y+\delta} f_Y(t)\,dt}$$

Divide the numerator and denominator by δ and note that

$$\frac{1}{\delta}\int_y^{y+\delta} f_Y(t)\,dt \xrightarrow[\delta\to 0]{}$$ (fundamental thm. of Calculus)

Similarly, since f_{XY} is cont., for any fixed $x\in I$,

$$\frac{1}{\delta}\int_y^{y+\delta} f_{XY}(x,t)\,dt \xrightarrow[\delta\to 0]{} f_{XY}(x,y).$$

It follows from the cont. of f_{XY} that

$$\frac{1}{\delta}\int_I\int_y^{y+\delta} f_{XY}(x,t)\,dt\,dx = \int_I\left[\frac{1}{\delta}\int_y^{y+\delta} f_{XY}(x,t)\,dt\right]dx$$

$$\xrightarrow[\delta\to 0]{} \int_I f_{XY}(x,y)\,dx.$$ ($\int\lim$ is not always $\lim\int$)

Therefore,

$$P(X\in I \mid y\le Y\le y+\delta) \xrightarrow[\delta\to 0]{} \frac{\int_I f_{XY}(x,y)\,dx}{f_Y(y)} = \int_I \frac{f_{XY}(x,y)}{f_Y(y)}\,dx.$$

This motivates the following definition for $y\in\mathbb{R}$:

<u>Def.</u> The conditional density of X given $Y=y$ for $y\in\mathbb{R}$ with $f_Y(y)>0$ is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$ (note the analogy with the discrete case)

Often one defines $f_{X|Y}(x|y)\equiv 0$ for y with $f_Y(y)=0$.

Note that $f_{X|Y}(x|y) \geq 0$ and that for y with $f_Y(y) > 0$,

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx = \int_{-\infty}^{\infty} \frac{f_{XY}(x,y)}{f_Y(y)}\,dx$$

$$= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \quad .$$

Thus $f_{X|Y}$ is indeed a pdf (in x) and therefore

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(s|y)\,ds$$

is the CDF of a continuous dist. with density $f_{X|Y}$. We call it the conditional dist. of X given $Y = y$, and $F_{X|Y}$ is the conditional CDF.

Note that

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} \frac{f_{XY}(s,y)}{f_Y(y)}\,ds$$

$$= \lim_{\delta \to 0} P(X \in \underbrace{(-\infty, x]}_{I \text{ is infinite but still works.}} \mid Y \in [y, y+\delta]),$$

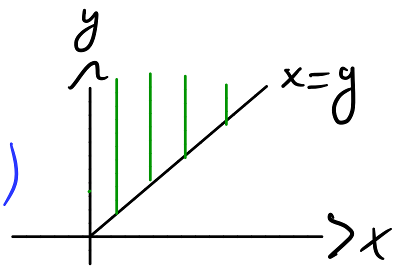
so this def. agrees with our intuition.

As expected, if X and Y are ind. (as well as jointly cont.) then

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

(makes sense)

Examples.



1) $f_{XY}(x,y) = \lambda^2 e^{-\lambda y} \cdot 1_{y > x \geq 0} \quad (\lambda > 0)$

$\Rightarrow f_X(x) = \lambda e^{-\lambda x} 1_{x \geq 0}, \quad f_Y(y) = \lambda^2 y e^{-\lambda y} 1_{y > 0}$

For $y > 0$,

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

$$= \begin{cases} \dfrac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = \dfrac{1}{y} & x \in [0,y) \\ 0 & x \notin [0,y) \end{cases}$$

$\Rightarrow X | Y = y \sim U(0,y)$.

2) Bivariate standard normal.

$$f_{Z|W}(z|w) = \frac{f_{ZW}(z,w)}{f_W(w)}$$

$$= \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(z^2 - 2\rho z w + w^2)}}{\frac{1}{\sqrt{2\pi}} e^{-w^2/2}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(z^2 - 2\rho z w + \rho^2 w^2)}$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(z - \rho w)^2}$$

$\Rightarrow Z | W = w \sim N(\rho w, \underbrace{1-\rho^2}_{\text{doesn't depend on } w!}) \quad (\rho \in (-1,1))$

We know how to sample from a univariate CDF F_X.
How can we sample from F_{XY}? Easy if ind. but otherwise?
It is easier to consider the discrete case first:

Suppose X and Y are discrete RVs with joint pmf p_{XY}. The identity

$$p_{XY}(x,y) = p_X(x)\, p_{Y|X}(y|x)$$

suggests we can sample from p_{XY} by first sampling $x \sim p_X$ (e.g., $Q_{F_X}(U)$) and then sample $y \sim p_{Y|X}(\cdot|x)$.
Similarly, in the cont. case,

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

implies that sampling $X \sim F_X$ $(Q_{F_X}(U))$ followed by sampling $Y \sim F_{Y|X=x}$ generates a sample from F_{XY}.
So, to sample from $f_{XY}(x,y) = \lambda^2 e^{-\lambda y} \cdot 1_{y>x\geq 0}$ we can sample from $Y \sim \Gamma(2,\lambda)$, then sample $X \sim U(0,y)$.

Similarly, we can sample from the bivariate normal f_{ZW} by first sampling $W \sim N(0,1)$ followed by drawing $Z \sim N(\rho W, 1-\rho^2)$.

Sums of RVs

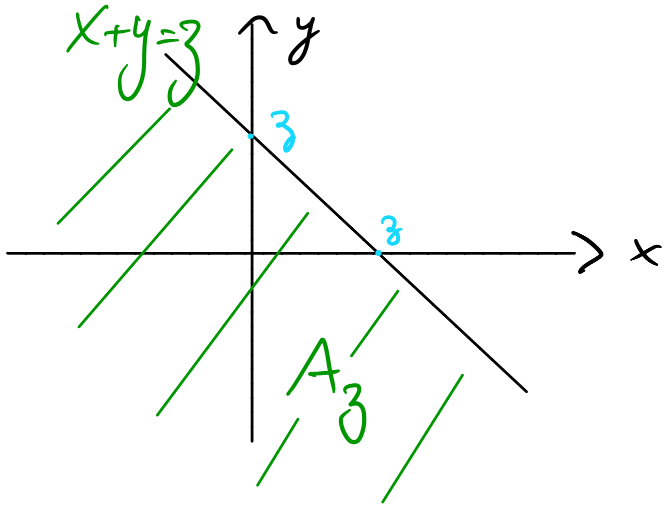
Recall: if X and Y are discrete, ind. $\mathbb{N}$-valued RVs, then $p_{X+Y}(z) = \sum_x p_X(x)\, p_Y(z-x)$ (convolution).

Let $Z = X+Y$ where X and Y have joint pdf f_{XY}.

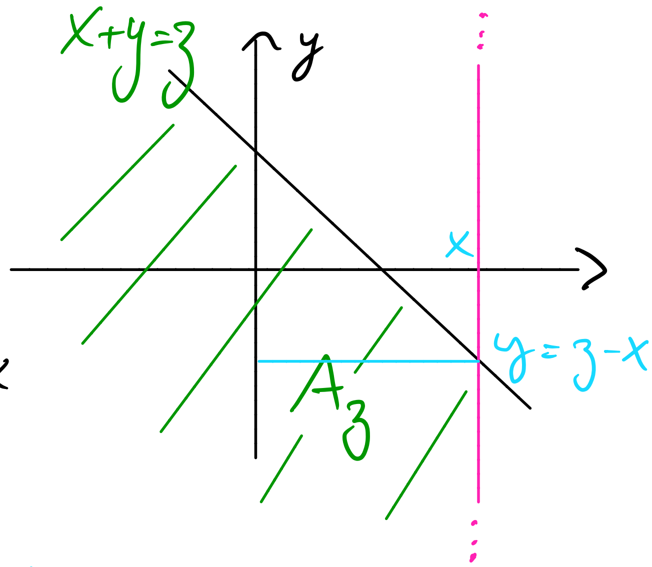
We want to find the dist. of Z again.

Consider the set

$$A_z = \{(x,y) \in \mathbb{R}^2 : x+y \le z\}:$$



$$F_Z(z) = P(X+Y \le z)$$
$$= P((X,Y) \in A_z)$$
$$= \iint_{A_z} f_{XY}(x,y)\, dx\, dy$$
$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\, dy\, dx$$



Change variable $v = y+x$

$$\Rightarrow F_Z(z) = \int_{x=-\infty}^{\infty} \int_{v=} f_{XY}(x, v-x)\, dv\, dx$$

Fubini
$$= \int_{v=-\infty}^{z} \left[\int_{x=-\infty}^{\infty} f_{XY}(x, v-x)\, dx \right] dv$$

$\Rightarrow$ Z is a cont. RV with density

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)\, dx .$$

If X and Y are ind. then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx ,$$ (compare with the discrete case)

The RHS is denoted as $f_X * f_Y$ which is the convolution of f_X and f_Y.

<u>Example.</u> $X \sim \Gamma(\alpha,\lambda)$ ind. of $Y \sim \Gamma(\beta,\lambda)$. $\alpha,\beta,\lambda>0$

$$\Rightarrow f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx \qquad z>0$$

$$= \int_{\boxed{}}^{\boxed{}} \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x} \frac{\lambda^\beta (z-x)^{\beta-1}}{\Gamma(\beta)} e^{-\lambda(z-x)}\, dx$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \cdot \int_0^z x^{\alpha-1}(z-x)^{\beta-1}\, dx.$$

$$\int_0^z x^{\alpha-1}(z-x)^{\beta-1}\, dx \overset{x=zt}{=} \int_0^1 (zt)^{\alpha-1}(z-zt)^{\beta-1} z\, dt$$

$$= z^{\alpha+\beta-1} \underbrace{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt}_{\beta(\alpha,\beta) \text{ Beta function}}$$

$$\Rightarrow f_{X+Y}(z) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)/\beta(\alpha,\beta)} \cdot z^{\alpha+\beta-1} \cdot e^{-\lambda z}$$

$\Rightarrow X+Y \sim \Gamma(\alpha+\beta, \lambda)$ (same rate, the shapes add up)

<u>Cor. 1</u>

$$\Gamma(\alpha+\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\beta(\alpha,\beta)}$$

or,

$$\beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Recall: $\exp(\lambda) \equiv \Gamma(1,\lambda)$

<u>Cor. 2</u> If $X_1, \dots, X_n$ are iid $\exp(\lambda)$ RVs then

$$\sum_1^n X_i = (\dots(((X_1 + X_2) + X_3) + X_4) + \dots + X_{n-1}) + X_n \sim$$

<u>Proof.</u> We saw that $X_i \sim \Gamma(1,\lambda)$ so by induction,

$$\sum_1^n X_i \sim \Gamma(n,\lambda).$$

In particular, when n is large, $\Gamma(n,\lambda)$ looks roughly normal. (why?)