Technical Difficulties so used the blackboard to go over Quiz 5

\[ f_{X,Y}(x,y) = \begin{cases} cx^2 & \text{for } 0 \leq y \leq 1-x^2 \\ 0 & \text{else} \end{cases} \]

Bivariate distribution \(2\) variables

Support is in two dimensions

All pdfs have \(2\) attributes

1) area = 1

2) non-negative

\[ \text{Support } S: \{ (x,y): 0 \leq y \leq 1-x^2 \} \]

\[
\iint_S f_{X,Y}(x,y) \, dy \, dx = 1
\]

\[
\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \, cy \, dx \, dy = \int_{-1}^{1} \int_{0}^{1-x^2} \, cy \, dx \, dy = \frac{4}{15} c = 1 \quad c = \frac{15}{4}
\]

Now back to normal lecture.

**Consequences of the definition of bivariate CDFs:**

1) **If** \((X, Y)\) **has the joint CDF** \(F_{X,Y}(x,y)\), **you can obtain the marginal CDF** \(F_X(x)\) **from the joint CDF** as \(F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y)\)
2) The joint pdf and joint CDF are related in a manner similar to their relationship in the univariate (one rv at a time) case:

If $(X, Y)$ have a joint pdf $f_{X,Y}(x, y)$ then
\[
F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(r, s) \, dr \, ds
\]

\[
= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(r, s) \, ds \, dr
\]

and $f_{X,Y}(x, y) = \frac{d^2}{dx \, dy} F_{X,Y}(x, y) = \frac{d^2}{dy \, dx} F_{X,Y}(x, y)$

at every $(x, y)$ where the partial derivatives exist.

3) If $(X, Y)$ have a discrete joint distribution with joint pmf $f_{X,Y}(x, y)$, then the marginal pmf $f_X(x)$ of $X$ is $f_X(x) = \sum_y f_{X,Y}(x, y)$

(and similarly for $f_Y(y)$)

$(X, Y)$: discrete You have $f_{X,Y}(x, y)$

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

The idea behind marginal distributions is that it's harder to visualize a joint (2-dimensional) distribution than it is to visualize each of its 1-dimensional marginal distributions.
4) If $(X, Y)$ have a continuous joint distribution with joint pdf $f_{X,Y}(x,y)$ the marginal pdf $f_{X,Y}(x,y)$, the marginal pdf $f_X(x)$ of $X$ is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dy \quad \text{(for all } -\infty < x < \infty)$$

Up: marginalizing over $Y$

and the marginal pdf $f_Y(y)$ of $Y$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dx \quad \text{(for all } -\infty < y < \infty)$$

Earlier example continued

$(X, Y)$ have joint pdf $f_{X,Y}(x,y) = \begin{cases} \frac{21}{8} x^2 & 0 \leq x^2 \leq 1 \\ 0 & \text{else} \end{cases}$

You can see from the sketch of the support $S_{X,Y}$ of $f_{X,Y}(x,y)$ that $-1 \leq x \leq 1$, so the support $S_X$ of $X$ is $(-1, 1)$

and its marginal pdf is $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dy$

$$= \int_{x^2}^{1} \frac{21}{8} x^2 \, dy = \frac{21}{8} x^2 (1-x^4)$$

so $f_X(x) = \begin{cases} \frac{21}{8} x^2 (1-x^4) & \text{for } -1 < x < 1 \\ 0 & \text{else} \end{cases}$

This is a weird pdf
(supposed to be symmetric and bimodal)
Similarly, the support $S_\mathcal{Y}$ of $\mathcal{Y}$ is $(0, 1)$ and its marginal pdf is

$$f_\mathcal{Y}(y) = \int_{-\infty}^{\infty} f_{\mathcal{X}\mathcal{Y}}(x, y) \, dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{2\lambda}{\pi} x^2 \gamma \, dx$$

$$= \left\{ \begin{array}{ll}
\frac{\lambda}{2} y^{3/2} & \text{for } 0 < y < 1 \\
0 & \text{else}
\end{array} \right.$$

3.5 $f_\mathcal{Y}(y)$

5) If you have the joint dist. $f_{\mathcal{X}\mathcal{Y}}(x, y)$ you can reconstruct the marginals $f_\mathcal{X}(x)$ and $f_\mathcal{Y}(y)$ but not the other way around:

If all you have is the marginals, in general they do not uniquely determine the joint.

\underline{Ex. Case 1:} $\mathcal{X}$ = # heads in $n$ tosses of fair coin 1 and independently $\mathcal{Y}$ = # heads in $n$ tosses of fair coin 2

$\mathcal{X} \sim \text{Binomial}(n, \frac{1}{2})$ so

$$f_\mathcal{X}(x) = \left\{ \begin{array}{ll}
\binom{n}{x} \left( \frac{1}{2} \right)^x \left( 1 - \frac{1}{2} \right)^{n-x} & x = 0, 1, \ldots, n \\
0 & \text{else}
\end{array} \right.$$ and $\mathcal{Y} \sim \text{Binomial}(n, \frac{1}{2})$ also so

$$f_\mathcal{Y}(y) = \left\{ \begin{array}{ll}
\binom{n}{y} \left( \frac{1}{2} \right)^y \left( 1 - \frac{1}{2} \right)^{n-y} & y = 0, 1, \ldots, n \\
0 & \text{else}
\end{array} \right.$$
Since $\overline{X}$ and $\overline{Y}$ are independent in case 1,
$$f_{\overline{X}\overline{Y}}(x, y) = f_{\overline{X}}(x) \cdot f_{\overline{Y}}(y)$$

If $A_i, B$ independent $\rightarrow P(A \cap B) = P(A) \cdot P(B)$

T/F statements

$$f_{\overline{X}\overline{Y}}(x, y) = (x)(y)(\frac{1}{2})^2 \quad \text{for } x = 0, 1, \ldots, n \quad \text{and} \quad y = 0, 1, \ldots, n$$

else

Case 2: $\overline{X} =$ # heads in $n$ tosses of fair coin 1

$\overline{Y} = X$

$X \sim \text{Binomial}(n, \frac{1}{2})$ and so is $\overline{Y}$ but then joint distribution (from $\overline{Y} = X$) is

$$f_{X\overline{Y}}(x, y) = \binom{n}{x}(\frac{1}{2})^n \quad \text{for } x = y = 0, \ldots, n$$

else

There is one situation in which the marginals do uniquely determine the joint: when $X$ and $\overline{Y}$ are independent.

**Def:** rvs $X$ and $\overline{Y}$ are independent if for every (non-empty) set $A$ and $B$ of real numbers

$$P(X \in A \text{ and } \overline{Y} \in B) = P(X \in A) \cdot P(\overline{Y} \in B)$$

**Consequences**

1. Immediately you get that if $X$ and $\overline{Y}$ are independent
\[F_{X \mid Y}(x \mid y) = P(X \leq x \text{ and } Y = y) = P(X \leq x) \cdot P(Y = y)\]

This is true if the converse is also true.

2) Differentiate this equation once with respect to \(y\) to get the result:

\(X, Y\) independent \iff \[F_{X \mid Y}(x \mid y) = f_X(x) \cdot f_Y(y)\]

Ex: Suppose that continuous r.v.s \(X\) and \(Y\) have joint pdf

\[f_{X \mid Y}(x \mid y) = \begin{cases} \kappa x^2 y^2 & \text{for } 0 \leq x^2 + y^2 \leq 1 \\ 0 & \text{else} \end{cases}\]

The support \(S_{X \mid Y}\) of \(f_{X \mid Y}\) is the region inside the unit circle.

You can evaluate the normalizing constant by computing \[\iint_{S_{X \mid Y}} k x^2 y^2 \, dx \, dy\] and setting it equal to 1.

\[1 = \int_{1}^{1} \int_{-1}^{1} k x^2 y^2 \, dy \, dx = \frac{\pi}{24} \quad \text{so} \quad k = \frac{24}{\pi}\]

Q: Are \(X\) and \(Y\) independent?

A: No, they can't be since the points \((X, Y)\) with positive density satisfy \(x^2 + y^2 \leq 1\). For any given value of \(y\) of \(Y\), the possible values of \(X\) depend on \(y\) and vice versa.
Ex: Continuous r.v $X$ and $Y$ have joint pdf $f_{X,Y}(x,y) = \begin{cases} ke^{-(x+2y)} & \text{for } x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{else} \end{cases}$

Q: Are $X$ and $Y$ independent?
A: Yes because $e^{-(x+2y)}$ factors into $(e^{-x})(e^{-2y})$ and the support $S_{X,Y}$ also factors: $(x \geq 0)(y \geq 0)$

Just choose $(k_x, k_y)$ such that $\int \int k_x e^{-(x+2y)} \, dx \, dy = 1$

$\int_0^\infty k_x e^{-x} \, dx = 1$, $\int_0^\infty k_y e^{-2y} \, dy = 1$ and $k = k_x \cdot k_y$

You get $k_x = 1$, $k_y = 2$, $k = 2$

Conditional probability distributions

Recalling that for two events $A$ and $B$, $P(B|A) = \frac{P(A \cap B)}{P(A)}$ (as long as $P(A) > 0$), we should be able to extend this idea to random variables.

Start with $X$ and $Y$ both discrete, so that we can talk about $P(Y = y | X = x)$

Def: If $X$ and $Y$ have a discrete joint probability distribution $f_{X,Y}(x,y)$ and $X$ has marginal p.m.f $f_X(x)$ then for each $x$ such that $f_X(x) > 0$ define

$f_{Y|X}(y|x) \equiv \frac{f_{X,Y}(x,y)}{f_X(x)}$ to be the conditional p.m.f of $Y$ given $X$

$= P(Y = y | X = x)$
Now let's do the analogous thing for continuous RVs.

**Def:** If $X$ and $Y$ have a continuous joint distribution with joint pdf $f_{X,Y}(x,y)$ and $X$ has continuous marginal pdf $f_X(x)$, then for each $x$ such that $f_X(x) > 0$, define $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ to be the conditional pdf of $Y$ given $X$.

Continuing an earlier ex.

$X, Y$ have joint pdf $f_{X,Y}(x,y) = \begin{cases} \frac{21}{4} x^2 y & \text{for } 0 \leq x^2 y \leq 1 \\ 0 & \text{else} \end{cases}$

Let's work out $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$.

Earlier we saw that $f_X(x) = \begin{cases} \frac{21}{8} x^2 (1-x^2) & \text{for } -1 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$

and $f_Y(y) = \begin{cases} \frac{7}{2} y^{3/2} & \text{for } 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$

Immediately, then, for all $x$ for which $f_X(x) > 0$, namely $-1 < x < 1$,

$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{21}{4} x^2 y & \text{for } 0 \leq x^2 y \leq 1 \\ \frac{21}{8} x^2 (1-x^2) & \text{else} \end{cases}$
This simplifies to
\[
 f_{\gamma 1 \bar{X}} (Y|X) = \begin{cases} 
 \frac{2\gamma}{1-x^2} & 0 \leq x^2 \leq \gamma \leq 1 \\
 0 & \text{else}
\end{cases}
\]

A few "slices" of this:

\[ f_{\gamma 1 \bar{X}} (Y|X) \]

\[ x = 0 \]
\[ 0 \]
\[ 1 \]
\[ y \]

\[ 0 \]
\[ 1 \]

And in the other direction
\[
 f_{\bar{X} 1 | Y} (X|Y) = \frac{f_{\bar{X} Y} (X,Y)}{f_Y(Y)}
\]
\[
 = \frac{2\gamma x^2 y}{\gamma^{5/2}} = \frac{3x^2}{2\gamma^{3/2}} = \frac{3}{2} x^2 y^{-3/2}
\]
\[
 = \begin{cases} 
 \frac{3}{2} x^2 y^{-3/2} & 0 \leq x^2 \leq \gamma \leq 1 \\
 0 & \text{else}
\end{cases}
\]

A few "slices" of this:

\[ f_{\bar{X} 1 | Y} (X|Y) \]

\[ y = 1 \]
\[ \frac{3}{2} x^2 \]
\[ -1 \]
\[ 1 \]

\[ y = \gamma \]
\[ 4.74 x^2 \]
\[ -1 \]
\[ 1 \]
When \( X \) and \( Z \) are continuous, computing \( f_{Z | X}(y | x) \) may seem to involve conditioning on the event \( \bar{X} = x \), which (as we saw earlier) has probability zero.

But that's not what's actually going on; strictly speaking \( f_{Z | X}(y | x) \) is a limit:

\[
f_{Z | X}(y | x^*) = \lim_{\varepsilon \to 0} \frac{d}{dy} P(\bar{X} \leq y \mid x^* - \varepsilon \leq X \leq x^* + \varepsilon)
\]

In other words, you take a little strip of \( x \) values of width \( \varepsilon \) around \( \bar{X} = x^* \), compute \( P(\bar{X} \leq y) \) \( X \) is in the strip), differentiate the result with respect to \( y \), and let \( \varepsilon \) go to 0.

Thus you can think of \( f_{Z | X}(y | x) \) as the conditional pdf of \( Z \) given that \( X \) is close to \( x \).

Constructing a joint pdf from marginals and conditionals, we know that as long as no divisions by 0 happen,

\[
f_{Z \mid X}(y | x) = \frac{f_{ZA}(x, y)}{f_X(x)} \quad (1)
\]

\[
f_{X \mid Z}(x | y) = \frac{f_{ZA}(x, y)}{f_Z(y)} \quad (2)
\]

Multiply \((1)\) by \( f_X(x) \) and equation \((2)\) by \( f_Z(y) \) to get

\[
f_{Z \mid X} = f_X(x) f_{Z \mid X}(y | x)
\]

\[
=f_Z(y) f_{X \mid Z}(x | y)
\]
\[ P(A \cap B) = P(A) P(B|A) \]
\[ = P(B) P(A|B) \]

So there are two ways to construct a joint pdf from a marginal pdf and a conditional pdf.

Case Study: Bayesian Statistical Analysis

A machine produces nuts \( \Theta \) and bolts \( \Theta \), and each nut paired with a particular bolt in the manufacturing process is supposed to fit snugly on the bolt.

Let's call a (nut, bolt) pair defective if the correct snug fit doesn't happen (e.g. bolt diameter either too big or too small, or nut diameter too small or too big).

Let \( \Theta \) = proportion of defective bolts if the machine were allowed to run for an indefinitely long period.

Since we can only observe the machine for a finite (short) time interval, \( \Theta \) is unknown.

To learn about \( \Theta \), we could take a random sample of (nut, bolt) pairs of size \( m \) ("with replacement") and count the \# of defectives in the sample (call this \( N \)).
Let $D_i = \begin{cases} 1 & \text{if (nut, bolt) pair is defective} \\ 0 & \text{else} \end{cases}$

$$(D_i; \theta) \sim \text{Bernoulli}(\theta) \quad \text{for } i = 1, \ldots, n$$

$${\sum}_{i=1}^{n} D_i = N$$

So the conditional pdf of rv $N$ is $f_{N|\theta}(n|1m, \theta)$

$$f_{N|\theta}(n|1m, \theta) = \begin{cases} \binom{m}{n} \theta^n (1-\theta)^{m-n} & \text{for } n = 0, 1, \ldots, m \\ 0 & \text{else} \end{cases}$$

Suppose that $m = 114, N = 3$

A reasonable estimate of $\theta$ would be $\hat{\theta} = \frac{N}{m} = \frac{3}{114} = 2.6\%$

but how much uncertainty do we have about $\theta$ on the basis of this dataset?

**Bayesian story**

- $\theta$ unknown continuous $\text{E}(0,1)$
- vector $D = (D_1, \ldots, D_m)$ data set

Probability: $P(\text{data|unknown}) =$ easy

$P(N|\theta)$ = (*)

$P(D|\theta) = P(\theta | D)$

**Bayes Theorem:** $P(\theta|D) = \frac{P(\theta) P(D|\theta)}{P(D)}$

$P(\theta|N) = \frac{P(\theta) P(N|\theta)}{P(N)} \propto \text{normalizing constant}$

- data info
- info about $\theta$
- external to dataset
- internal to dataset