Previously found a form of Bayes' Theorem and the TPL in terms of density functions (marginal and conditional).

Now consider the case where two rvs X and Y are independent. Then

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

if $f_X(x) \neq 0$

Note: You must be careful if you want to use $f_{Y|X}(y|x) = f_Y(y)$

for independence of $X, Y,$ since this is valid only if $f_X(x) \neq 0$.

Now summarize Bayes' Theorem for rvs:

① If X and Y are both discrete, use

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

with $A = \{X = x\}, B = \{Y = y\}$ for $x \in R_X$, $y \in R_Y$

this can be written in terms of pmfs as

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

for all $x \in R_X$, $y \in R_Y$ where $p_X(x) \neq 0$, $p_Y(y) \neq 0$

conditional pmf of X given $Y=y$, defined as

$$p_{X|Y}(x|y) = P(X = x|Y = y)$$

(2) If X continuous, Y discrete, use

$$f_X(x|B) = \frac{P(B|X = x) f_X(x)}{P(B)}$$

with $B = \{Y = y\}$. This can be written as

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

for $x \in \mathbb{R}$, $y \in R_Y$, $f_X(x) \neq 0$, $p_Y(y) \neq 0$

③ X continuous, Y continuous

Use

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

$\forall x, y \in \mathbb{R}$, where $f_X(x) \neq 0$, $f_Y(y) \neq 0$

Comment for HW 10:
Some definitions

- The <u>nth moment</u> of a rv X is

$$E[X^n], \quad n = 1, 2, \ldots$$

- The <u>nth central moment</u> of X is

$$E\left[(X - \overline{X})^n\right], \quad n = 2, 3, \ldots$$

- The <u>ln-th moment</u> of rvs X, Y is

$$E\left[X^l Y^n\right], \quad l = 0, 1, \ldots; \quad n = 1, 2, \ldots$$

- The <u>ln-th central moment</u> of X, Y is

$$E\left[\left(X-\overline{X}\right)^{l}\left(Y-\overline{Y}\right)^{n}\right],$$

$$l=0,1,\ldots;\ n=1,2,\ldots$$

The Laws of Large Numbers

Consider a rv X that you wish to model. For example, X might be the steady-state temperature of a system. You might take n measurements $X^{(1)},\ldots,X^{(n)}$ of X. Assume that these measurements are independent and all have the same distribution, which is referred to as being "independent and identically distributed (iid)".

Now let

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}$$

Y_n is a <u>sample mean</u> of X computed from samples $X^{(1)}, \ldots, X^{(n)}$.

Sample means are used to estimate expected values, so Y_n is an estimate of $E[X]$.

Is it a good estimate? Does it get closer to $E[X]$ as the number of samples increases?

The Laws of Large Numbers (LLNs) of which there are two, address this issue. These laws both say that, if $\mu_X < \infty$ and $\sigma_X < \infty$, then

$$Y_n \to E[X] \quad \text{as} \quad n \to \infty,$$

"in some sense". The Strong LLN states that

$Y_n \to E[X]$ with probability one, which means that

unless an event of probability zero happens the sequence of sample means Y_n converges to the true mean $E[X]$.
The Weak LLN states the same result but in a "weaker" sense than with prob. 1.

The LLNs can also be used to show that relative frequency probabilities will converge to the measure P as the number of trials converges to infinity, as follows:

Consider a random experiment and an event A in that experiment for which you would like to know the prob P(A). The relative frequency approach estimates P(A) by running the experiment n times and computing the number of times A occurs divided by n.

Is this a good estimate of P(A)? What happens as the number of trials converges to ∞?

Define a rv X_n as

$$X_n = \begin{cases} 1, & \text{if A occurs in trial n} \\ 0, & \text{otherwise} \end{cases}$$

Then compute the sample mean

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

By the SLLN, if the trials are independent,

$$Y_n \to E[X_i] \quad \text{as } n \to \infty$$

(with prob. 1), where

$$E[X_i] = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0)$$

$$= P(A)$$

for any $i = 1, \ldots, n$.

This means that the relative frequency estimate converges to the true probability $P(A)$.