Summary:

6.4 Conditional Distributions: Discrete case
6.5 Conditional Distributions: Continuous case
6.7 Joint Distribution of Functions of R.V.

6.4 Conditional Distribution: Discrete case

\[ p_{X|Y}(x|y) = P(X=x | Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} \]

is called the conditional probability mass of \(X\) given that \(Y = y\), and

\[ F_{X|Y}(x|y) : \mathbb{R} \to [0,1] \text{ defined by} \]

\[ F_{X|Y}(x|y) = P(X \leq x | Y=y) = \sum_{a \leq x} p_{X|Y}(a|y) \]

is called the conditional distribution function of \(X\) given that \(Y = y\).
Remark. If $X, Y$ are independent and $y < y' < Y$, then
\[
F_{X|Y}(x|y') = F_X(x) \quad \forall x \leq y'
\]
\[
F_{X|Y}(x|y) = F_X(x) \quad \forall x \leq y
\]

Applications: $X_1, X_2$ independent, Poisson r.v. with parameters $\lambda_1, \lambda_2$ then
\[
X_i | X_1 + X_2 = \nu = \text{B}(\nu, \frac{\lambda_1}{\lambda_1 + \lambda_2})
\]

Multinomial distribution:
\[
P(X_1 = u_1, X_2 = u_2, \ldots, X_k = u_k) = \frac{\nu!}{u_1! \cdot u_2! \cdots u_k!} \cdot \nu_1^{u_1} \nu_2^{u_2} \cdots \nu_k^{u_k}
\]

\[
u_i \geq 0 \quad \sum_{i=1}^{k} u_i = \nu \quad \text{for the proper } u_i
\]

For any $1 \leq k \leq \nu$, any $u_1, u_2, \ldots, u_k$ such that $\sum_{j=1}^{k} u_j = k \leq \nu$,
\[
P(X_1 = u_1, \ldots, X_k = u_k | X_{k+1} = u_{k+1}, \ldots, X_{\nu} = u_{\nu})
\]
\[
= \frac{(\nu - k)!}{u_1! \cdots u_k!} \cdot \left( \frac{\nu_1}{\nu} \right)^{u_1} \cdots \left( \frac{\nu_k}{\nu} \right)^{u_k}
\]

For any $u_1, u_2, \ldots, u_k$ with $u_1 + u_2 + \cdots + u_k = k$. 

§ 6.5 Conditional Distributions: continuous case

**Def.** If $X, Y$ are jointly continuous and $f_Y(y) > 0$ then

$$f_{X|Y}(x|y) : \mathbb{R} \rightarrow \mathbb{R}$$

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

is called the conditional probability density of $X$ given

that $Y = y$, and

$$F_{X|Y}(x|y) : \mathbb{R} \rightarrow [0,1]$$

$$F_{X|Y}(a|y) = \frac{a}{y} f_{X|Y}(x|y) \, dx$$

is the conditional distribution of $X$ given $Y = y$.

**Remark.** The definition comes from a limit argument:

$$P(X \leq a \mid Y = y) = \lim_{dy \to 0} P(X \leq a \mid y < Y < y + dy)$$
\[
\lim_{\Delta y \to 0} \frac{\mathbb{P}(x \leq a, y < Y < y + \Delta y)}{\Delta y - \mathbb{P}(y < Y < dy)} = \lim_{\Delta y \to 0} \frac{\int f(x,y) \, dx \, dy}{\Delta y}
\]

\[
\text{L'Hospital's Rule:}
\]

\[
= \frac{a}{\int f(x,y) \, dx} / f_y(y)
\]

\[
\int_a^\infty \frac{f(x,y) \, dx}{f_y(y)}
\]

Taking the \( a \) derivative, \( f_x|_y \, (a/\gamma) = \frac{f(a,y)}{f_y(y)} \)

Applications: The Bivariate Normal Distribution

A joint distribution of \( X, Y \) with joint probability density:

\[
f(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} \right] \right)
\]

where \( \sigma_x, \sigma_y > 0 \) and \(-1 < \rho < 1\).

Then \( X, Y \) are both normal distributions with means \( \mu_x, \mu_y \), and variances \( \sigma_x^2 \), respectively \( \sigma_y^2 \), and \( X/Y = Y \) is a normal distribution with mean \( \mu_x + \frac{\sigma_x}{\sigma_y} (y-\mu_y) \) and variance \( \sigma_x^2 (1-\rho^2) \), see textbook.
§ 6.7 Joint distribution of functions of r.v.

Let $X_1, X_2 : S \rightarrow \mathbb{R}$ be discrete r.v. with joint probability mass $\mathbb{P}_{X_1, X_2}$.

Let $g = (g_1, g_2) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ bijective and

$$(Y_1, Y_2) = g(X_1, X_2), \text{ i.e. } Y_1 = g_1(X_1, X_2), Y_2 = g_2(X_1, X_2)$$

Then the joint probability mass for $Y_1$ and $Y_2$ is:

$$P(Y_1 = y_1, Y_2 = y_2) = P(g_1(X_1, X_2) = y_1, g_2(X_1, X_2) = y_2)$$

$$= P(X_1 = x_1, X_2 = x_2)$$

$$= \mathbb{P}_{X_1, X_2}(x_1, x_2)$$

where $g(X_1, X_2) = (Y_1, Y_2)$ or $(X_1, X_2) = g^{-1}(Y_1, Y_2)$

$Y_1, X_1, X_2$ are jointly continuous with joint probability density $f_{X_1, X_2}$, and we also assume $g$ to be a local diffeomorphism, then $Y_1, Y_2$ are jointly continuous with joint probability density.
\[ f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2) \left| \frac{\partial (y_1,y_2)}{\partial (x_1,x_2)} \right|^{-1} \]

where \((y_1,y_2) = g(x_1,x_2)\) or \((x_1,x_2) = g^{-1}(y_1,y_2)\)

**Proof**

\[
F_{Y_1,Y_2}(y_1,y_2) = \mathbb{P}(g_1(x_1,x_2) \leq y_1, g_2(x_1,x_2) \leq y_2)
\]

\[= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1,x_2) \, dx_1 \, dx_2 \]

Taking the \(\frac{\partial^2}{\partial y_1 \partial y_2}\) derivative gives the result.
Applications: See example 7b

Generalizations to n random variables:

\( X_1, X_2, \ldots, X_n \) discrete r.v.

\[ g: \mathbb{R}^n \to \mathbb{R}^n \text{ bijective} \]

\( (Y_1, Y_2, \ldots, Y_n) = g(X_1, X_2, \ldots, X_n) \) then

\[ P(Y_1 = y_1, \ldots, Y_n = y_n) = P(X_1 = x_1, \ldots, X_n = x_n) \text{ where} \]

\[ (X_1, \ldots, X_n) = g^{-1}(Y_1, \ldots, Y_n) \]

\( X_1, \ldots, X_n \) jointly continuous r.v. with joint probability density \( f_{X_1, X_2, \ldots, X_n} \),

\[ g: \mathbb{R}^n \to \mathbb{R}^n \text{ bijective and local diffeomorphism}, \]

then \( (Y_1, Y_2, \ldots, Y_n) = g(X_1, X_2, \ldots, X_n) \) are also jointly continuous with joint probability density:

\[ f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n) = f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \left| \frac{\partial g}{\partial x_1, x_2, \ldots, x_n} \right|^{-1} \]