Joint Distributions, Discrete Case

In the following, X and Y are **discrete** random variables.

1. **Joint distribution (joint p.m.f.):**

   - **Definition:** $f(x,y) = P(X = x, Y = y)$
   - **Properties:** (1) $f(x,y) \geq 0$, (2) $\sum_{x,y} f(x,y) = 1$
   - **Representation:** The most natural representation of a joint discrete distribution is as a **distribution matrix**, with rows and columns indexed by x and y, and the xy-entry being $f(x,y)$. This is analogous to the representation of ordinary discrete distributions as a single-row table. As in the one-dimensional case, the entries in a distribution matrix must be nonnegative and add up to 1.

2. **Marginal distributions:** The distributions of X and Y, when considered separately.

   - **Definition:**
     - $f_X(x) = P(X = x) = \sum_y f(x,y)$
     - $f_Y(y) = P(Y = y) = \sum_x f(x,y)$
   - **Connection with distribution matrix:** The marginal distributions $f_X(x)$ and $f_Y(y)$ can be obtained from the distribution matrix as the row sums and column sums of the entries. These sums can be entered in the "margins" of the matrix as an additional column and row.
   - **Expectation and variance:** μ_X, μ_Y, σ_X^2, σ_Y^2 denote the (ordinary) expectations and variances of X and Y, computed as usual: $\mu_X = \sum_x x f_X(x)$, etc.

3. **Computations with joint distributions:**

   - **Probabilities:** Probabilities involving X and Y (e.g., $P(X + Y = 3)$ or $P(X \geq Y)$ can be computed by adding up the corresponding entries in the distribution matrix: More formally, for any set R of points in the xy-plane, $P((X,Y) \in R)) = \sum_{(x,y)\in R} f(x,y)$.
   - **Expectation of a function of X and Y (e.g., $u(x,y) = xy$):** $E(u(X,Y)) = \sum_{x,y} u(x,y)f(x,y)$. This formula can also be used to compute expectation and variance of the marginal distributions directly from the joint distribution, without first computing the marginal distribution. For example, $E(X) = \sum_{x,y} x f(x,y)$.

4. **Covariance and correlation:**

   - **Definitions:** $\operatorname{Cov}(X,Y) = E(XY) - E(X)E(Y) = E((X - \mu_X)(Y - \mu_Y))$ (Covariance of X and Y), $\rho = \rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}$ (Correlation of X and Y)
   - **Properties:** $|\operatorname{Cov}(X,Y)| \leq \sigma_X \sigma_Y$, $-1 \leq \rho(X,Y) \leq 1$
   - **Relation to variance:** $\operatorname{Var}(X) = \operatorname{Cov}(X,X)$
   - **Variance of a sum:** $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$ (Note the analogy of the latter formula to the identity $(a+b)^2 = a^2 + b^2 + 2ab$; the covariance acts like a "mixed term" in the expansion of $\operatorname{Var}(X+Y)$.)

5. **Independence of random variables:**

   - **Definition:** X and Y are called independent if the joint p.m.f. is the product of the individual p.m.f.'s: i.e., if $f(x,y) = f_X(x)f_Y(y)$ **for all values of** x **and** y.
   - **Properties of independent random variables:**
     If X and Y are independent, then:
     - **The expectation of the product of X and Y is the product of the individual expectations:** $E(XY) = E(X)E(Y)$. More generally, this product formula holds for any expectation of a function X times a function of Y. For example, $E(X^2Y^3) = E(X^2)E(Y^3)$.
     - **The product formula holds for probabilities of the form P(some condition on X, some condition on Y) (where the comma denotes "and"):** For example, $P(X \le 2, Y \le 3) = P(X \le 2)P(Y \le 3)$.
     - **The covariance and correlation of X and Y are 0:** $\mathrm{Cov}(X,Y) = 0$, $\rho(X,Y) = 0$.
     - **The variance of the sum of X and Y is the sum of the individual variances:** $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
     - **The moment-generating function of the sum of X and Y is the product of the individual moment-generating functions:** $M_{X+Y}(t) = M_X(t)M_Y(t)$. (Note that it is the **sum** $X+Y$, not the product XY, which has this property.)

6. **Conditional distributions:**

   - **Definitions:**
     - **conditional distribution (p.m.f.) of X given that $Y = y$:**
       $g(x|y) = P(X = x|Y = y) = \frac{f(x,y)}{f_Y(y)}$
     - **conditional distribution (p.m.f.) of Y given that $X = x$:**
       $h(y|x) = P(Y = y|X = x) = \frac{f(x,y)}{f_X(x)}$
   - **Connection with distribution matrix:** Conditional distributions are the distributions obtained by fixing a row or column in the matrix and rescaling the entries in that row or column so that they again add up to 1. For example, $h(y|2)$, the conditional distribution of Y given that $X = 2$, is the distribution given by the entries in row 2 of the matrix, rescaled by dividing by the row sum (namely, $f_X(2)$): $h(y|2) = f(2,y)/f_X(2)$.
   - **Conditional expectations and variance:** Conditional expectations, variances, etc., are defined and computed as usual, but with conditional distributions in place of ordinary distributions:
     - $E(X|y) = E(X|Y = y) = \sum_x x g(x|y)$
     - $E(X^2|y) = E(X^2|Y = y) = \sum_x x^2 g(x|y)$
     - $\mathrm{Var}(X|y) = \mathrm{Var}(X|Y = y) = E(X^2|y) - E(X|y)^2$

     More generally, for any condition (such as $Y > 0$), the expectation of X given this condition is defined as

     - $E(X|\,\text{condition}) = \sum_x x P(X = x|\,\text{condition})$

     and can be computed by starting out with the usual formula for the expectation, but restricting to those terms that satisfy the condition.

Joint Distributions, Continuous Case

In the following, X and Y are **continuous** random variables. Most of the concepts and formulas below are analogous to those for the discrete case, with integrals replacing sums. The principal difference between continuous lies in the definition of the p.d.f./p.m.f. $f(x, y)$: The formula $f(x, y) = P(X = x, Y = y)$ is no longer valid, and there is no simple and direct way to obtain $f(x, y)$ from X and Y.

1. **Joint continuous distributions:**
   - **Joint density (joint p.d.f.):** A function $f(x, y)$ satisfying (i) $f(x, y) \geq 0$, (ii) $\iint f(x, y)dxdy = 1$. Usually, $f(x, y)$ will be given by an explicit formula, along with a **range** (a region in the xy-plane) on which this formula holds. *In the general formulas below, if a range of integration is not explicitly given, the integrals are to be taken over the range in which the density function is defined.*
   - **Uniform joint distribution:** An important special type of joint density is one that is constant over a given range (a region in the xy-plane), and 0 outside outside this range, the constant being the reciprocal of the area of of the range. This is analogous to the concept of an ordinary (one-variable) uniform density $f(x)$ over an interval I, which is constant (and equal to the reciprocal of the length of I) inside the interval, and 0 outside it.

2. **Marginal distributions:** The ordinary distributions of X and Y, when considered separately. The corresponding (one-variable) densities are denoted by f_X (or f_1) and f_Y (or f_2), and obtained by integrating the joint density $f(x, y)$ over the "other" variable:

$$f_X(x) = \int f(x, y)dy, \quad f_Y(y) = \int f(x, y)dx.$$

3. **Computations with joint distributions:**
   - **Probabilities:**
     Given a region R in the xy-plane the probability that (X, Y) falls into this region is given by the double integral of $f(x, y)$ over this region. For example, $P(X + Y \leq 1)$ is given by an integral of the form $\iint_R f(x, y)dxdy$, where R consists of the part of the range of f in which $x + y \leq 1$.
   - **Expectation of a function of X and Y (e.g., $u(x, y) = xy$):**
     $E(u(X, Y)) = \iint u(x, y)f(x, y)dxdy$

4. **Covariance and correlation:**
   The formulas and definitions are the same as in the discrete case.
   - **Definitions:** $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y) = E((X - \mu_X)(Y - \mu_Y))$ (Covariance of X and Y), $\rho = \rho(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$ (Correlation of X and Y)
   - **Properties:** $|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y$, $-1 \leq \rho(X, Y) \leq 1$
   - **Relation to variance:** $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$
   - **Variance of a sum:** $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$

5. **Independence of random variables:** Same as in the discrete case:
   - **Definition:** X and Y are called independent if the joint p.d.f. is the product of the individual p.d.f.'s: i.e., if $f(x, y) = f_X(x)f_Y(y)$ for all x, y.

- **Properties of independent random variables:**
  If X and Y are independent, then:
  - **The expectation of the product of X and Y is the product of the individual expectations:** $E(XY) = E(X)E(Y)$. More generally, this product formula holds for any expectation of a function X times a function of Y. For example, $E(X^2Y^3) = E(X^2)E(Y^3)$.
  - **The product formula holds for probabilities of the form P(some condition on X, some condition on Y) (where the comma denotes "and"):** For example, $P(X \leq 2, Y \leq 3) = P(X \leq 2)P(Y \leq 3)$.
  - **The covariance and correlation of X and Y are 0:** $\mathrm{Cov}(X, Y) = 0$, $\rho(X, Y) = 0$.
  - **The variance of the sum of X and Y is the sum of the individual variances:** $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
  - **The moment-generating function of the sum of X and Y is the product of the individual moment-generating functions:** $M_{X+Y}(t) = M_X(t)M_Y(t)$.

6. **Conditional distributions:** Same as in the discrete case, with integrals in place of sums:

- **Definitions:**
  - **conditional density of X given that $Y = y$:**
    $g(x|y) = \frac{f(x,y)}{f_Y(y)}$
  - **conditional density of Y given that $X = x$:**
    $h(y|x) = \frac{f(x,y)}{f_X(x)}$
- **Conditional expectations and variance:** Conditional expectations, variances, etc., are defined and computed as usual, but with conditional distributions in place of ordinary distributions. For example:
  - $E(X|Y = 1) = E(X|Y = 1) = \int xg(x|1)dx$
  - $E(X^2|Y = 1) = E(X^2|Y = 1) = \int x^2g(x|1)dx$