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) $\sum\limits_{x,y} f(x,y) = 1$
   - **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:**
     - $f_X(x) = P(X = x) = \sum\limits_{y} f(x,y)$
     - $f_Y(y) = P(Y = y) = \sum\limits_{x} f(x,y)$
   - **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: $\mu_X = \sum\limits_{x} x f_X(x)$, 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, $P((X,Y) \in R)) = \sum\limits_{(x,y) \in R} f(x,y)$.
   - **Expectation of a function of X and Y (e.g., $u(x,y) = xy$):** $E(u(X,Y)) = \sum\limits_{x,y} u(x,y) f(x,y)$. 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, $E(X) = \sum\limits_{x,y} x f(x,y)$.

4. **Covariance and correlation:**

   - **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:** $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$ (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 $\mathrm{Var}(X + Y)$.)

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, $P(X \leq 2, Y \leq 3) = P(X \leq 2)P(Y \leq 3)$.
  - **The covariance and correlation of X and Y are** 0: $\text{Cov}(X, Y) = 0$, $\rho(X, Y) = 0$.
  - **The variance of the sum of X and Y is the sum of the individual variances:** $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
  - **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:
  - $E(X|y) = E(X|Y = y) = \sum_x xg(x|y)$
  - $E(X^2|y) = E(X^2|Y = y) = \sum_x x^2 g(x|y)$
  - $\text{Var}(X|y) = \text{Var}(X|Y = y) = E(X^2|y) - E(X|y)^2$

  More generally, for any condition (such as $Y > 0$), the expectation of X given this condition is defined as
  - $E(X|\text{condition}) = \sum_x xP(X = x|\text{condition})$

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