Useful Probability Properties

Jonathan Scarlett

January 8, 2020

- **General notation:**
  - If X if a discrete random variable, its probability mass function (PMF) is written as P_X (i.e., $P_X(x) = \mathbb{P}[X = x]$).
  - If X is a continuous random variable, its probability density function (PDF) is written as f_X (e.g., $\mathbb{P}[a \le X \le b] = \int_a^b f_X(x)dx$). **For now (and most of the course), let's assume all random variables are discrete.**
  - We will usually use upper case for a random variable (e.g., X) and the corresponding lower-case letter for a specific value (e.g., x).
- **Expectation:**
  - Definition: $\mathbb{E}[X] = \sum_x P_X(x)x$
  - Average of function: $\mathbb{E}[f(X)] = \sum_x P_X(x)f(x)$ for deterministic f
  - Average of scaled RV: $\mathbb{E}[cX] = c\mathbb{E}[X]$ for constant c
  - Average of sum: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ regardless of whether or not X and Y are independent
  - Average of product: If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
  - Indicator function: If $\mathbf{1}\{A\}$ denotes the indicator function (equaling 1 if the event A holds and 0 otherwise), then $\mathbb{E}[\mathbf{1}\{A\}] = \mathbb{P}[A]$
- **Conditioning:**
  - Definition: $P_{Y|X}(y|x) = \frac{P_{XY}(x,y)}{P_X(x)}$
  - Law of total probability: For an event A and RV X, we have $\mathbb{P}[A] = \sum_x P_X(x)\mathbb{P}[A|X = x]$
  - Law of total expectation: (AKA tower property) $\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y|X]\big]$, where the outer expectation is over X and the inner one is over Y (given X)

- Bayes' rule: $\mathbb{P}[A|B] = \frac{\mathbb{P}[A]\mathbb{P}[B|A]}{\mathbb{P}[B]}$

- **Independence:**
  - Definition: $P_{XY}(x,y) = P_X(x)P_Y(y)$ for all x, y
  - Equivalent definition 1: $P_{Y|X}(y|x) = P_Y(y)$ for all x, y
  - Equivalent definition 2: $P_{X|Y}(x|y) = P_X(x)$ for all x, y
  - Analogous definitions for conditional independence: (i) $P_{XY|Z}(x,y|z) = P_{X|Z}(x|z)P_{Y|Z}(y|z)$ for all x, y, z; (ii) $P_{Y|XZ}(y|x,z) = P_{Y|Z}(y|z)$ for all x, y, z; (iii) $P_{X|YZ}(x|y,z) = P_{X|Z}(x|z)$ for all x, y, z. We use the terminology "X and Y are conditionally independent given Z".
  - Functions: If X and Y are independent, then so are $f(X)$ and $g(Y)$ for deterministic f, g
  - Conditional vs. unconditional: The statements "X and Y are independent" and "X and Y are conditionally independent given Z" can be very different:
    - Example 1: If X and Y are independent and $Z = X + Y$, then X and Y are certainly not conditionally independent given Z
    - Example 2: If U and V are independent and $X = Z + U$, $Y = Z + V$, then U and V are conditionally independent given Z, but dependent due to the common reliance on Z
  - Joint independence of a collection $X_1, \dots, X_n$ of random variables can be defined as $P_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^n P_{X_i}(x_i)$.
    - Note: Pairwise independence does not necessarily imply joint independence

- **Variance:**
  - Definition: $\mathrm{Var}[X] = \mathbb{E}\big[(X-\mu)^2\big]$, where μ is the mean of X
  - Equivalent definition: $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2$
  - Scaling: $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$ for constant c
  - Variance of sum: If X and Y are independent, then $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$. (More generally, $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$.)
  - Covariance: $\mathrm{Cov}[X,Y] = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$ where $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$
  - Law of total variance: (Not needed, see Wikipedia if interested)

- **Other:**
  - Marginal distribution: $P_X(x) = \sum_y P_{XY}(x,y)$ and similarly $P_Y(y) = \sum_x P_{XY}(x,y)$

- Union vs. intersection: $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$
- Union bound: (AKA Boole's inequality) $\mathbb{P}\big[\bigcup_{i=1}^{N} A_i\big] \leq \sum_{i=1}^{N} \mathbb{P}[A_i]$
- Law of large numbers: If $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) with mean μ, then $\mathbb{P}\big[\big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\big| > \epsilon\big] \to 0$ as $n \to \infty$ for arbitrarily small $\epsilon > 0$

- **Properties of logarithms (not related to probability):**
  - $\log xy = \log x + \log y$
  - $\log \frac{1}{x} = -\log x$
  - $\log \frac{y}{x} = \log y - \log x$
  - $\log x^c = c \log x$
  - $\log_a x = \frac{\log_b x}{\log_b a}$
  - $\log_e x \leq x - 1$ with equality if and only if $x = 1$

- **Very basic calculus (not related to probability):**
  - $\frac{d}{dx} x^c = c x^{c-1}$
  - $\frac{d}{dx} e^{cx} = c e^{cx}$
  - $\frac{d}{dx} \ln x = \frac{1}{x}$
  - Chain rule: $\frac{d}{dx} f(g(x)) = f'(g(x))\, g'(x)$
  - Product rule: $\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$
  - Quotient rule: $\frac{d}{dx}(f(x)/g(x)) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$