# CSC 665: Probability review

Chicheng Zhang

August 29, 2019

## 1 Probability review

1. **Probability.** Denote by $\mathbb{P}(A)$ the probability of event $A$; (e.g. throwing a die, $A = \{$ number 6 is up $\}$, $\mathbb{P}(A) = 1/6$.)

   Probability satisfies additivity: if $A \cap B = \emptyset$, i.e. they are mutually exclusive, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. It also satisfies subadditivity: for general $A, B$, $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

   Events $A$ and $B$ are called independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

   Union bound: $\mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) \leq \mathbb{P}(A_1) + \ldots + \mathbb{P}(A_n)$.

2. **Expectation.** For a random variable $X$, denote by its expectation $\mathbb{E}[X]$. Specifically, if $X$ takes value in a discrete set $S$, with probability mass function $p$, then

   $$\mathbb{E}[X] \triangleq \sum_{x \in S} x \cdot p(x);$$

   If $X$ is continuous and has probability density function $p$, then

   $$\mathbb{E}[X] \triangleq \int_{\mathbb{R}} x \cdot p(x) dx.$$

3. **Indicator function.** Denote by indicator function

   $$\mathbf{1}(A) \triangleq \begin{cases} 1 & A \text{ is true}, \\ 0 & A \text{ is false}. \end{cases}$$

   As $\mathbf{1}(A)$ only takes values 0 and 1, we immediately get from the definition of expectation that

   $$\mathbb{E}\mathbf{1}(A) = 1 \times \mathbb{P}(A) + 0 \times \mathbb{P}(\bar{A}) = \mathbb{P}(A).$$

4. **Linearity of expectation.** Suppose $X$, $Y$ are two (possibly dependent) random variables. Then, $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

   Is $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$? This is not true in general (consider $(X, Y)$ as having the joint distribution of taking $(-1, -1)$ and $(+1, +1)$ with probability 0.5.) However, if $X$ and $Y$ are independent, then

   $$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

   is true. Furthermore, if $X$ and $Y$ are independent, then for any functions $f$ and $g$,

   $$\mathbb{E}f(X)g(Y) = \mathbb{E}[f(X)]\mathbb{E}[g(Y)].$$

5. **Variance.** Recall that the variance of a random variable $X$ (with mean $\mu$) is defined as:

$$\mathrm{Var}(X) \triangleq \mathbb{E}(X-\mu)^2.$$

By linearity of expectation,

$$\mathbb{E}(X-\mu)^2 = \mathbb{E}X^2 - \mathbb{E}2X \cdot \mu + \mu^2 = \mathbb{E}X^2 - \mu^2.$$

How does $\mathrm{Var}(X+Y)$ relate to $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$? Again, there is no equation relationship in general. However a notable fact is that if $X$ and $Y$ are independent, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. This is because,

$$\mathrm{Var}(X+Y) = \mathbb{E}(X+Y-\mathbb{E}X-\mathbb{E}Y)^2 = \mathbb{E}(X-\mathbb{E}X)^2 + \mathbb{E}(Y-\mathbb{E}Y)^2 + 2\mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

6. **Jensen's Inequality.** Recall that a convex function $f$ is one that for all $x_1$, $x_2$ in $\mathbb{R}$, and $t \in [0,1]$,

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

Useful facts:

(a) A twice differentiable function is convex if and only if its second derivative is always nonnegative. (This provides a practical way to check convexity.)

(b) If $f$ is differentiable, then for any $x$,$y$, $f(y) \geq f(x) + f'(x)(y-x)$. That is, $f$ is always above its first-order approximation. (For twice differentiable $f$, this is a direct consequence of Taylor's Theorem: $f(y) = f(x) + f'(x)(y-x) + \frac{f''(\xi)}{2}(y-x)^2$ for some $\xi$ between $x$ and $y$.)

**Theorem 1.** *Suppose $f$ is a convex function, and $X$ is a random variable. Then*

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

*Proof.* We only show the inequality when $f$ is differentiable. Denote by $\mu \triangleq \mathbb{E}X$. Observe that for all $x$, $f(x) \geq f(\mu) + f'(\mu)(x-\mu)$. Taking expectation on both sides, we get that $\mathbb{E}f(X) \geq f(\mu) + \mathbb{E}f'(\mu)(X-\mu) = f(\mu) = f(\mathbb{E}X)$. □

7. **Markov's inequality:** a positive random variable with bounded mean should not take large values too often.

**Theorem 2** (Markov's Inequality)**.** *Suppose $X$ is a nonnegative random variable. Then for any $a > 0$, $\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}X}{a}$.*

*Proof.* Observe that for any positive $x$, $x \geq a\mathbf{1}(x \geq a)$. Therefore,

$$\mathbb{E}X \geq \mathbb{E}a\mathbf{1}(X \geq a) = a\mathbb{P}(X \geq a).$$

The proof is concluded by dividing both sides by $a$. □

8. **Chebyshev's Inequality:** a random variable with a bounded variance should not deviate from its mean too often.

**Theorem 3** (Chebyshev's Inequality)**.** *Suppose $X$ is a random variable with mean $\mu$ and variance $v > 0$. Then for any $b > 0$, $\mathbb{P}[|X-\mu| \geq b] \leq \frac{v}{b^2}$.*

*Proof.* Applying Markov's Inequality to the random variable $Y = (X-\mu)^2$ and $a = b^2$, we get

$$\mathbb{P}((X-\mu)^2 \geq b^2) \leq \frac{\mathbb{E}Y}{b^2}.$$

The proof is concluded by noting that the event $\{|X-\mu| \geq b\}$ is the same as the event $\{(X-\mu)^2 \geq b^2\}$, and the fact that $\mathbb{E}Y = v$. □