Lecture #21: Tail Bounds
Key Concepts

Basic Bounds

Theorem (Basic Inequality)**.** *Let Ω be a* **finite** *sample space with probability distribution P : $\Omega \to \mathbb{R}$, let $X : \Omega \to \mathbb{R}$ be a random variable, and let $h : \mathbb{R} \to \mathbb{R}$ be a total function such that*

$$h(x) \geq 0 \quad \textit{for all } x \in \mathbb{R}.$$

Then, for every real number a such that $a > 0$,

$$\mathsf{P}(h(X) \geq a) \leq \frac{\mathsf{E}[h(X)]}{a}.$$

Corollary (Markov's Inequality)**.** *Let Ω be a* **finite** *sample space with probability distribution P : $\Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$ be a random variable. Then, for every* **positive** *real number a,*

$$\mathsf{P}(|X| \geq a) \leq \frac{\mathsf{E}[|X|]}{a}.$$

Variance and Standard Deviation

Definition. Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$. Then the ***variance*** of X, with respect to P, is

$$\mathsf{var}(X) = \sum_{\mu \in \Omega} (X(\mu) - \mathsf{E}[X])^2 \times \mathsf{P}(\mu)$$

and the ***standard deviation*** of X, denoted $\sigma(X)$, is $\sqrt{\mathsf{var}(X)}$.

Theorem. *Let Ω be a* **finite** *sample space, let P : $\Omega \to \mathbb{R}$ be a probability distribution for Ω, and let X be a random variable. Then X^2 is also a random variable, and*

$$\mathsf{var}(X) = \mathsf{E}[X^2] - \mathsf{E}[X]^2.$$

It is *not* generally true that var$(X+Y) = $ var$(X) + $ var(Y) for a pair of random variables X and Y — so the following result is useful (and not trivial):

Theorem. *Let Ω be a* **finite** *sample space with probability distribution P : $\Omega \rightarrow \mathbb{R}$ and let $X_1, X_2, \ldots, X_n : \Omega \rightarrow \mathbb{R}$ be random variables (for some positive integer n). If $X_1, X_2, \ldots, X_n$ are* **pairwise independent** *then*

$$\text{var}(X_1 + X_2 + \cdots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \cdots + \text{var}(X_n).$$

More Bounds

Theorem (Chebyshev's Inequality)**.** *Let Ω be a* **finite** *sample space with probability distribution P : $\Omega \rightarrow \mathbb{R}$, let X be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then*

$$\mathsf{P}(|X| \geq a) \leq \frac{\mathsf{E}[X^2]}{a^2}.$$

and

$$\mathsf{P}(|X - \mathsf{E}[X]| \geq a) \leq \frac{\text{var}(X)}{a^2}.$$

Theorem (Cantelli's Inequality)**.** *Let Σ be a* **finite** *sample space with probability distribution P : $\Omega \rightarrow \mathbb{R}$, let $X : \Omega \rightarrow \mathbb{R}$ be a random variable, and let $a \in \mathbb{R}$ such that $a > 0$. Then*

$$\mathsf{P}(X - \mathsf{E}[X] \geq a) \leq \frac{\text{var}(X)}{a^2 + \text{var}(X)}.$$

Cantelli's Inequality is sometimes called the "One-Sided Chebyshev's Inequality".

Let e be ***Euler's number*** — so that e is approximately (but not exactly) 2.71828.

Theorem (Chernoff Bound)**.** *Let Ω be a finite sample space and let* P : $\Omega \rightarrow \mathbb{R}$ *be a probability distribution for Ω. Suppose that $X_1, X_2, \ldots, X_n$ are mutually independent random variables such that $X_i : \Omega \rightarrow \{0, 1\}$ for $1 \leq i \leq n$, and suppose that* $\mathsf{P}(X_i = 1) = p_i$ *for every integer i such that $1 \leq i \leq n$ — so that $p_1, p_2, \ldots, p_n$ are real numbers such that $0 \leq p_i \leq 1$ for $1 \leq i \leq n$. Let $X = X_1 + X_2 + \cdots + X_n$, and let* $\mu = \mathsf{E}[X] = p_1 + p_2 + \cdots + p_n > 0$.

(a) For every real number δ such that $\delta > 0$,

$$\mathsf{P}(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right)^{\mu}.$$

(b) For every real number δ such that $0 < \delta < 1$,

$$\mathsf{P}(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{((1-\delta)^{1-\delta}} \right)^{\mu},$$

Corollary. *Let Ω be a a finite sample space with probability distribution P : $\Omega \to \mathbb{R}$. Suppose that $X_1, X_2, \ldots, X_n$ are mutually independent random variables such that $X_i : \Omega \to \{0,1\}$ for $1 \le i \le n$, and suppose that $\mathsf{P}(X_i = 1) = p_i$, for $1 \le i \le n$ — so that $p_1, p_2, \ldots, p_n$ are real numbers such that $0 \le p_i \le 1$ for $1 \le i \le n$. Let $X = X_1 + X_2 + \cdots + X_n$ and let $\mu = \mathsf{E}[X] = p_1 + p_2 + \cdots + p_n > 0$.*

(a) If $0 < \delta \le 1$ then $\mathsf{P}(X \ge (1+\delta)\mu) \le e^{-\frac{\delta^2\mu}{3}}$.

(b) For all positive real numbers γ and δ, if $\delta \ge \gamma$ then $\mathsf{P}(X \ge (1+\delta)\mu) \le \left(e^{\gamma-(1+\gamma)\ln(1+\gamma)}\right)^{\mu}$.

(c) If $0 < \delta < 1$ then

$$\mathsf{P}(X \le (1-\delta)\mu) \le e^{-\frac{\delta^2\mu}{2}}.$$

Corollary. *Let Ω be a a finite sample space with probability distribution P : $\Omega \to \mathbb{R}$. Suppose that $X_1, X_2, \ldots, X_n$ are mutually independent random variables such that $X_i : \Omega \to \{0,1\}$ for $1 \le i \le n$, and suppose that $\mathsf{P}(X_i = 1) = p$, for $1 \le i \le n$ — for a real number p such that $0 < p < 1$. Let $X = X_1 + X_2 + \cdots + X_n$.*

(a) If $\delta > 0$ then

$$\mathsf{P}(X \ge (1+\delta)pn) \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{pn}.$$

(b) If $0 < \delta < 1$ then

$$\mathsf{P}(X \le (1-\delta)pn) \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{pn}.$$

(c) If $0 \le \delta \le 1$ then $\mathsf{P}(X \ge (1+\delta)pn) \le e^{-\frac{\delta^2 pn}{3}}$.

(d) For all positive real numbers γ and δ, if $\delta \ge \gamma$ then $\mathsf{P}(X \ge (1+\delta)pn) \le \left(e^{\gamma-(1+\gamma)\ln(1+\gamma)}\right)^{pn}$.

(e) If $0 < \delta < 1$ then $\mathsf{P}(X \le (1-\delta)pn) \le e^{-\frac{\delta^2 pn}{2}}$.