Lecture #21: Tail Bounds
Proofs of Claims

Proof of the Basic Inequality

Theorem 3 (Basic Inequality)**.** *Let Ω be a* **finite** *sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, 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 $a > 0$,

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

Proof. Suppose, to obtain a contradiction, that

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

Then, since $a > 0$ it follows (by multiplying both sides of the inequality by a) that

$$a \times \mathsf{P}(h(X) \geq a) > \mathsf{E}[h(X)]. \tag{1}$$

Now (since Ω is a finite, and one can reorder the terms in a finite sum without changing its value)

$$\begin{aligned}
\mathsf{E}[h(X)] &= \sum_{\mu \in \Omega} h(X(\mu)) \times \mathsf{P}(\mu) \\
&= \sum_{\substack{\mu \in \Omega \\ h(X(\mu)) < a}} h(X(\mu)) \times \mathsf{P}(\mu) + \sum_{\substack{\mu \in \Omega \\ h(X(\mu)) \geq a}} h(X(\mu)) \times \mathsf{P}(\mu) \qquad \text{(splitting the sum)} \\
&\geq \sum_{\substack{\mu \in \Omega \\ h(X(\mu)) < a}} 0 \times \mathsf{P}(\mu) + \sum_{\substack{\mu \in \Omega \\ h(\mu) \geq a}} h(X(\mu)) \times \mathsf{P}(\mu) \\
&\qquad \text{(since } h(X(\mu)) \geq 0 \text{ and } \mathsf{P}(\mu) \geq 0 \text{ for every outcome } \mu \in \Omega)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\substack{\mu\in\Omega \\ h(X(\mu))\geq a}} h(X(\mu)) \times \mathsf{P}(\mu) \\
&\geq \sum_{\substack{\mu\in\Omega \\ h(X(\mu))\geq a}} a \times \mathsf{P}(\mu) && \text{(again, since } \mathsf{P}(\mu) \geq 0 \text{ for every outcome } \mu \in \Omega\text{)} \\
&= a \times \sum_{\substack{\mu\in\Omega \\ h(X(\mu))\geq a}} \mathsf{P}(\mu) \\
&= a \times \mathsf{P}(h(X) \geq a) \\
&> \mathsf{E}[X] && \text{(by the inequality at line (1), above).}
\end{aligned}
$$

Thus $\mathsf{E}[h(X)] > \mathsf{E}[h(X)]$ — which is impossible, since a real number cannot be strictly greater than itself. Since a ***assumption*** has been obtained, the assumption must be false — and

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

as claimed. □

Proof of Markov's Inequality

Corollary 4 (Markov's Inequality)**.** *Let* Ω *be a* **finite** *sample space with probability distribution* $\mathsf{P} : \Omega \to \mathbb{R}$*, and let* $X : \Omega \to \mathbb{R}$ *be a random variable. Then, for every* **positive** *real number* a*,*

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

Proof. This follows immediately from Theorem 3: In particular, Markov's Inequality follows by an application of the Basic Inequality, using the function $h : \mathbb{R} \to \mathbb{R}$ such that $h(x) = |x|$ for every real number x. □

Alternative Form of Variance

Theorem 5. *Let* Ω *be a* **finite** *sample space, let* $\mathsf{P} : \Omega \to \mathbb{R}$ *be a probability distribution for* Ω*, and let* X *be a random variable. Then* X^2 *is also a random variable, and*

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

Proof. Since Ω is a finite sample space,

$$\begin{aligned}
\mathsf{var}(X) &= \sum_{\mu\in\Omega}(X(\mu)-\mathsf{E}[X])^2 \times \mathsf{P}(\mu) \\
&= \sum_{\mu\in\Omega}(X(\mu)^2 - 2\mathsf{E}[X]\times X(\mu) + \mathsf{E}[X]^2)\times \mathsf{P}(\mu) \\
&= \left(\sum_{\mu\in\Omega} X(\mu)^2\times\mathsf{P}(\mu)\right) - \left(\sum_{\mu\in\Omega} 2\mathsf{E}[X]\times X(\mu)\times\mathsf{P}(\mu)\right) + \left(\sum_{\mu\in\Omega}\mathsf{E}[X]^2\times\mathsf{P}(\mu)\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(reordering terms)} \\
&= \mathsf{E}[X^2] - 2\mathsf{E}[X]\times\left(\sum_{\mu\in\Omega}X(\mu)\times\mathsf{P}(\mu)\right) + \mathsf{E}[X]^2\times\sum_{\mu\in\Omega}\mathsf{P}(\mu) \\
&= \mathsf{E}[X^2] - 2\mathsf{E}[X]\times\mathsf{E}[X] + \mathsf{E}[X]^2\times 1 \\
&= \mathsf{E}[X^2] - 2\mathsf{E}[X]^2 + \mathsf{E}[X]^2 \\
&= \mathsf{E}[X^2] - \mathsf{E}[X]^2,
\end{aligned}$$

as claimed. ☐

Using Pairwise Independence

Theorem 6. *Let Ω be a* **finite** *sample space with probability distribution $\mathsf{P}:\Omega\to\mathbb{R}$ and let $X_1, X_2, \ldots, X_n : \Omega\to\mathbb{R}$ be random variables (for some positive integer n). If $X_1, X_2, \ldots, X_n$ are* **pairwise independent** *then*

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

Recall, from Lecture #20, that the expected values of random variables are not generally “multiplicative”. As the following claim states, they *are* multiplicative when the random variables are independent.

Claim. *Let Ω be a* **finite** *sample space with probability distribution $\mathsf{P}:\Omega\to\mathbb{R}$, and let $X, Y : \Omega\to\mathbb{R}$ be* **independent** *random variables (with respect to P). Then*

$$\mathsf{E}[X\times Y] = \mathsf{E}[X]\times\mathsf{E}[Y].$$

Proof. Since the sample space Ω is finite there is a finite set of values

$$V_X = \{\alpha_1, \alpha_2, \ldots \alpha_k\} \subseteq \mathbb{R}$$

such that, if

$$S_i = \{\mu \in \Omega \mid X(\mu) = \alpha_i\}$$

for $1 \le i \le k$, then $S_i \ne \emptyset$ for $1 \le i \le k$ and

$$S_1 \cup S_2 \cup \cdots \cup S_k = \Omega.$$

Assuming (as the above notation may suggest) that $\alpha_1, \alpha_2, \ldots, \alpha_k$ are distinct (so that $|V_X| = k$), $S_i \cap S_j = \emptyset$, as well, for $1 \le i, j \le k$ such that $i \ne j$.

Now

$$\begin{aligned}
\mathsf{E}[X] &= \sum_{\mu \in \Omega} X(\mu) \times \mathsf{P}(\mu) \\
&= \sum_{i=1}^{k} \left(\sum_{\mu \in S_i} X(\mu) \times \mathsf{P}(\mu) \right) && \text{(reordering terms)} \\
&= \sum_{i=1}^{k} \left(\sum_{\mu \in S_i} \alpha_i \times \mathsf{P}(\mu) \right) && \text{(since } X(\mu) = \alpha_i \text{ for every outcome } \mu \in S_i\text{)} \\
&= \sum_{i=1}^{k} \left(\alpha_i \times \sum_{\mu \in S_i} \mathsf{P}(\mu) \right) \\
&= \sum_{i=1}^{k} \alpha_i \times \mathsf{P}(S_i) \\
&= \sum_{i=1}^{k} \alpha_i \times \mathsf{P}(X = \alpha_i).
\end{aligned}$$

Similarly, there is a finite set of values

$$V_Y = \{\beta_1, \beta_2, \ldots \beta_\ell\} \subseteq \mathbb{R}$$

such that, if

$$T_i = \{\mu \in \Omega \mid Y(\mu) = \beta_i\}$$

for $1 \le i \le \ell$, then $T_i \ne \emptyset$ for $1 \le i \le \ell$ and

$$T_1 \cup T_2 \cup \cdots \cup T_\ell = \Omega.$$

Assuming (as the above notation may suggest) that $\beta_1, \beta_2, \ldots, \beta_\ell$ are distinct (so that $|V_Y| = \ell$), $T_i \cap T_j = \emptyset$, as well, for $1 \le i, j \le \ell$ such that $i \ne j$. Repeating the above argument (replacing X with Y and the set V_X with V_Y) that

$$\mathsf{E}[Y] = \sum_{j=1}^{\ell} \beta_j \times \mathsf{P}(Y = \beta_j).$$

Note that, for every outcome $\mu \in \Omega$, there exists *exactly one* pair of integers i and j such that $\mu \in S_i \cap T_j$. Thus

$$
\begin{aligned}
\mathsf{E}[X \times Y] &= \sum_{\mu \in \Omega} (X \times Y)(\mu) \times \mathsf{P}(\mu) \\
&= \sum_{\mu \in \Omega} X(\mu) \times Y(\mu) \times \mathsf{P}(\mu) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} \left(\sum_{\mu \in S_i \cap T_j} X(\mu) \times Y(\mu) \times \mathsf{P}(\mu) \right) && \text{(reordering terms)} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} \left(\sum_{\mu \in S_i \cap T_j} \alpha_i \times \beta_j \times \mathsf{P}(\mu) \right) && \text{(by the definitions of } S_i \text{ and } T_j\text{)} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} \left(\alpha_i \times \beta_j \times \sum_{\mu \in S_i \cap T_j} \mathsf{P}(\mu) \right) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} (\alpha_i \times \beta_j \times \mathsf{P}(\mu \in S_i \cap T_j)) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} (\alpha_i \times \beta_j \times \mathsf{P}(\mu \in S_i \text{ and } \mu \in S_j)) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} (\alpha_i \times \beta_j \times \mathsf{P}(X = \alpha_i \text{ and } Y = \beta_j)) && \text{(by the definitions of } S_i \text{ and } T_j\text{)} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} \big(\alpha_i \times \beta_j \times \mathsf{P}(X = \alpha_i) \times \mathsf{P}(Y = \beta_{)}\big) \\
& && \text{(since } X \text{ and } Y \text{ are } \textbf{\textit{independent}} \text{ random variables)} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{\ell} (\alpha_i \times \mathsf{P}(X = \alpha_i)) \times (\beta_j \times \mathsf{P}(Y = \beta_j)) \\
&= \left(\sum_{i=1}^{k} \alpha_i \times \mathsf{P}(X = \alpha_i) \right) \times \left(\sum_{j=1}^{\ell} \beta_j \times \mathsf{P}(Y = \beta_j) \right) \\
& && \text{(reordering terms, once again)} \\
&= \mathsf{E}[X] \times \mathsf{E}[Y],
\end{aligned}
$$

as claimed. $\square$

Proof of Theorem 6. Let

$$X = X_1 + X_2 + \cdots + X_n.$$

Then, by Theorem 5, above,

$$\begin{aligned}
\mathsf{var}(X) &= \mathsf{E}[X^2] - \mathsf{E}[X]^2 \\
&= \mathsf{E}\left[\left(\textstyle\sum_{i=1}^n X_i\right)^2\right] - \left(\mathsf{E}\left[\textstyle\sum_{i=1}^n X_i\right]\right)^2 \\
&= \mathsf{E}\left[\textstyle\sum_{i=1}^n \sum_{j=1}^n X_i \times X_j\right] - \left(\mathsf{E}\left[\textstyle\sum_{i=1}^n X_i\right]\right)^2 \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathsf{E}[X_i \times X_j] - \left(\sum_{i=1}^n \mathsf{E}[X_i]\right)^2 && \text{(by Linearity of Expectation)} \\
&= \sum_{i=1}^n \sum_{j=a}^n \mathsf{E}[X_i \times X_j] - \sum_{i=1}^n \sum_{j=1}^n \mathsf{E}[X_i] \times \mathsf{E}[X_j] && \text{(reordering terms)} \\
&= \sum_{i=1}^n \left(\mathsf{E}[X_i^2] - \mathsf{E}[X_i]^2\right) + \sum_{\substack{1 \le i,j \le n \\ i \ne j}} \left(\mathsf{E}[X_i \times X_j] - \mathsf{E}[X_i] \times \mathsf{E}[X_j]\right) \\
& && \text{(reordering terms, again)} \\
&= \sum_{i=1}^n \left(\mathsf{E}[X_i^2] - \mathsf{E}[X_i]^2\right) + \sum_{\substack{1 \le i,j \le n \\ i \ne j}} \left(\mathsf{E}[X_i] \times \mathsf{E}[X_j] - \mathsf{E}[X_i] \times \mathsf{E}[j]\right) \\
& && \text{(by the above claim, since } X_i \text{ and } X_j \text{ are } \textbf{\textit{independent}} \text{ if } i \ne j) \\
&= \sum_{i=1}^n \left(\mathsf{E}[X_i^2] - \mathsf{E}[X_i]^2\right) + \sum_{\substack{1 \le i,j \le n \\ i \ne j}} 0 \\
&= \sum_{i=1}^n \left(\mathsf{E}[X_i^2] - \mathsf{E}[X_i]^2\right) \\
&= \sum_{i=1}^n \mathsf{var}(X_i).
\end{aligned}$$

That is,

$$\mathsf{var}(X_1 + X_2 + \cdots + X_n) = \mathsf{var}(X_1) + \mathsf{var}(X_2) + \cdots + \mathsf{var}(X_n),$$

as claimed. □

Chebyshev's Inequality and Cantelli's Inequality

The lecture notes also include results that can be applied, to used the expected values of variances of random variables to establish tail bounds, namely, ***Chebyshev's Inequality*** and ***Cantelli's Inequality*** (stated as Theorem 7 and Theorem 8, respectively). These will be considered in the lecture presentation and the tutorial exercise for this topic.

Proof of the Chernoff Bound

Theorem 9 (The Chernoff Bound)**.** *Let* Ω *be a finite sample space with probability distribution* Pr : $\Omega \to \mathbb{R}$*. Suppose that* $X_1, X_2, \ldots, X_n$ *are mutually independent random variables such that* $X_i : \Omega \to \{0, 1\}$ *for* $1 \leq i \leq n$*, and suppose that* $\mathsf{P}(X_i = 1) = p$ *for every integer* i *such that* $1 \leq i \leq n$*, for a real number* p *such that* $0 \leq p \leq 1$*. Let* $X = X_1 + X_2 + \cdots + X_n$*. Then, for every real number* θ *such that* $0 \leq \theta \leq 1$*,*

$$\mathsf{P}(X \geq (1+\theta)pn) \leq e^{-\frac{\theta^2}{3}pn}.$$

Sketch of Proof. Let t be any positive real number. Then, since X is a random variable, e^{tX} is a non-negative random variable — and

$$\mathsf{P}(X \geq (1+\theta)pn) = \mathsf{P}(e^{tX} \geq e^{t(1+\theta)pn}).$$

Now, since $X = X_1 + X_2 + \cdots + X_n$,

$$\mathsf{E}[e^{tX}] = \mathsf{E}[e^{tx_1} \times e^{tx_2} \times \cdots \times e^{tx_n}],$$

and, since the random variables $x_1, x_2, \ldots, x_n$ are mutually independent, so are the random variable $e^{tx_1}, e^{tx_2}, \ldots, e^{tx_n}$. This can be used to show, by induction on n, that

$$\mathsf{E}[e^{tX}] = \mathsf{E}[e^{tx_1} \times e^{tx_2} \times \cdots \times e^{tx_n}] = \prod_{i=1}^{n} \mathsf{E}[e^{tx_i}]. \tag{2}$$

Since the random variable x_i only assumes values 0 and 1, with probabilities p and $1-p$ respectively, the random variable e^{tx_i} only assumes values $e^0 = 1$ and e^t, with probabilities p and $1-p$ respectively, so that

$$\mathsf{E}[e^{tx_i}] = p \cdot 1 + (1-p) \cdot e^t = 1 - p(e^t - 1). \tag{3}$$

It now follows by the equations at lines (2) and (3) that

$$\mathsf{E}[e^{tX}] = \prod_{i=1}^{n} \mathsf{E}[e^{tx_i}] = (1 + p(e^t - 1))^n.$$

Now recall, by Markov's Inequality, that

$$\mathsf{P}(e^{tX} \geq k \cdot \mathsf{E}[e^{tX}]) \leq \tfrac{1}{k}$$

for any positive real number k. In particular, this is true when $k = e^{t(1+\theta)pn} \cdot \mathsf{E}[e^{tX}]^{-1}$, so that

$$\begin{aligned}\mathsf{P}(e^{tX} \geq e^{t(1+\theta)pn}) &\leq \frac{\mathsf{E}[e^{tX}]}{e^{t(1+\theta)pn}} \\ &= \frac{(1+p(e^t-1))^n}{e^{t(1+\theta)pn}}.\end{aligned}$$

A consideration of the Taylor expansion of the function $f(x) = e^x$ can be used to establish that $1 + x \leq e^x$ for every positive real number x, so that $(1+x)^n \leq e^{xn}$ for every positive real number x as well. Since t is a positive real number $e^t - 1 > 0$ as well, so that

$$(1 + p(e^t-1))^n \leq \left(e^{p(e^t-1)}\right)^n = e^{pn(e^t-1)}$$

and it now follows that

$$\mathsf{P}(X \geq (1+\theta)pn) = \mathsf{P}(e^{tX} \geq e^{t(1+\theta)pn}) \leq \frac{e^{pn(e^t-1)}}{e^{t(1+\theta)pn}}.$$

Now let $t = \ln(1+\theta)$ — which is a positive real number, since $\theta > 0$. Then

$$\begin{aligned}\mathsf{P}(X \geq (1+\theta)pn) &\leq \frac{e^{pn(e^t-1)}}{e^{t(1+\theta)pn}} \\ &= \frac{e^{\theta pn}}{e^{(1+\theta)pn\ln(1+\theta)}} \\ &= e^{pn(\theta-(1+\theta)\ln(1+\theta))} \\ &= e^{pnf(\theta)}\end{aligned}$$

for the function f such that $f(x) = x - (1+x)\ln(1+x)$ for every positive real number x. Now notice that $f'(x) = -\ln(1+x)$, $f''(x) = -(1+x)^{-1}$, $f^{(3)}(x) = (1+x)^{-2}$, and $f^{(\ell)}(x) = (-1)^{\ell+1} \cdot (\ell-2)!(1+x)^{\ell-1}$ for every integer ℓ such that $\ell \geq 4$. A consideration of a Taylor expansion for f (at 0) confirms that if θ is a real number such that $0 \leq \theta \leq 1$ then

$$\begin{aligned}f(\theta) &= \sum_{i\geq 2}(-1)^{i-1} \cdot \tfrac{1}{i\cdot(i-1)}\theta^i \\ &\leq -\tfrac{1}{2}\theta^2 + \tfrac{1}{6}\theta^3 \\ &\leq -\tfrac{1}{2}\theta^2 + \tfrac{1}{6}\theta^2 \\ &= -\tfrac{1}{3}\theta^2.\end{aligned}$$

Thus $e^{pnf(\theta)} \leq e^{-\frac{\theta^2}{3}pn}$, and it now follows that

$$\mathsf{P}(X \geq (1+\theta)pn) \leq e^{-\frac{\theta^2}{3}pn},$$

as claimed. $\square$