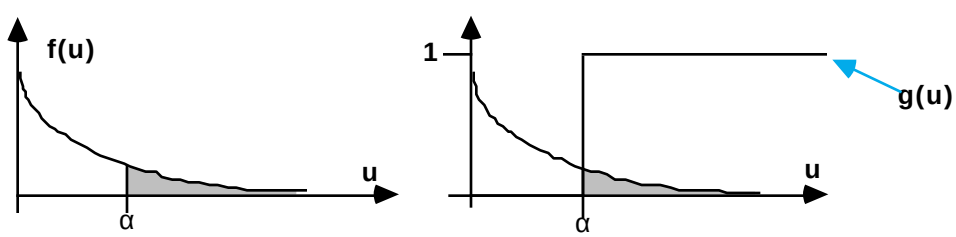
Chapter 6 Limit Theorems

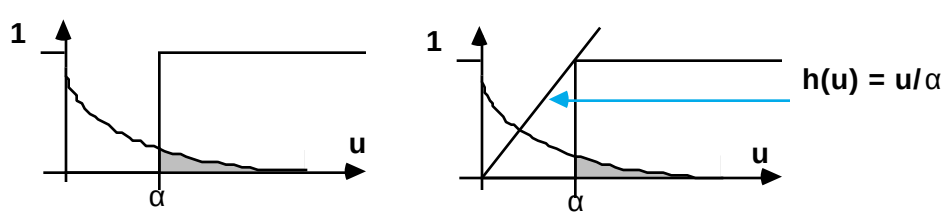
- In this chapter, we consider various bounds on probabilities that don't depend on specific forms of CDFs/pdfs/pmfs
- Relative frequencies and large numbers of trials
- Laws of large numbers
- Asymptotic distributions of random variables
- $P\{\mathbf{X} > u\} = 1 - F_{\mathbf{X}}(u) \to 0$ as $u \to \infty$
- Given the pdf/pmf/CDF, we can find $P\{\mathbf{X} > 5\}$ exactly
- What can be said about $P\{\mathbf{X} > 5\}$ when we don't know the probabilistic description?
- How small is $P\{\mathbf{X} > 5\}$?
- **Fundamental idea:** If $g(x) \le h(x)$ for all x (a,b), then $\int_a^b g(x)dx \le \int_a^b h(x)dx$
- *Markov's Inequality* is an upper bound on $P\{\mathbf{X} > \alpha\}$ for non-negative random variables with finite mean μ
- **Markov's Inequality:** If $\mathbf{X}$ is a non-negative random variable with finite mean μ, then, for any $\alpha > 0$,

$$P\{\mathbf{X} > \alpha\} \le \frac{\mu}{\alpha}$$

- The bound is uninteresting for $\alpha \le \mu$
- Bound $\to 0$ as $\alpha \to \infty$
- **Proof:** $P\{\mathbf{X} > \alpha\} = \int_a^\infty f(u)du = \int_0^\infty g(u)f(u)du$ where $g(u) = \begin{cases} 1, & u > \alpha, \\ 0, & \text{elsewhere.} \end{cases}$



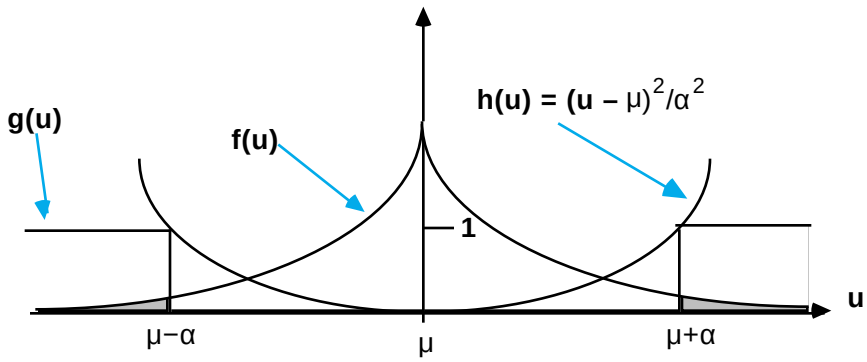
- $P\{\mathbf{X} > \alpha\} = \int_0^\infty g(u)f(u)du \le \int_0^\infty h(u)f(u)du$ where $h(u) = u/\alpha$



- $P\{\mathbf{X} > \alpha\} \le \int_0^\infty (u/\alpha)f(u)du = \frac{\mu}{\alpha}$
- $P\{\mathbf{X} > \alpha\} \le \mu/\alpha$

- Because the bound is so general, it can be applied when very little is known about the distribution
- Because the bound is so generally applicable, it is often quite loose
- **Example:** $P\{\mathbf{X} > 10\} \le 10^{-1}$ for *all* random variables with average value 1
- If $\mathbf{X}$ is an exponential random variable with parameter 1, i.e., $\mu = 1$, then $P\{\mathbf{X} > 10\} = \exp(-10) \approx 4.54\times10^{-5} << 10^{-1}$
- But, do we *know* the distribution of $\mathbf{X}$?
- For $\lambda > 0$, the **Chernoff bound** uses $\exp \lambda(u-\alpha)$ as an upper bound on the step function g
- $P\{\mathbf{X} > \alpha\} \le \int_0^\infty \exp \lambda(u-\alpha)f(u)du = \exp(-\lambda\alpha)\, E[\exp(\lambda\mathbf{X})]$
- For exponential RV, $P\{\mathbf{X} > \alpha\} \le \exp(-\lambda\alpha)/(1-\lambda)$
- **Example:** A bit transmitted over a data link is received incorrectly with probability p. To improve reliability, each data bit is sent n times. If there are more 1's than 0's among the n bits received , the receiver decides that a 1 was sent. Otherwise, it decides that a 0 was sent
- Assume that n transmissions are independent of each other
- $\mathbf{X}$ = number of incorrectly received bits (among n)
- $\mathbf{X}$ is a binomial random variable with parameters (n,p)
- $P\{\text{receiver decision incorrect}\} = P\{\mathbf{X} > n/2\} = \sum_{i > n/2} P\{X = i\} = \sum_{i > n/2} \binom{n}{i} p^i(1-p)^{n-i}$
- We can compute this given any n and p. But how do we solve the problem in reverse? That is, if we know p, say $p = 10^{-2}$, what should n be so that $P\{\mathbf{X} > n/2\} < 10^{-5}$?
- The Chernoff bound is handy here. $P\{\mathbf{X} > \alpha\} \le \exp(-\lambda\alpha)\, E[\exp(\lambda\mathbf{X})]$
- $E[\exp(\lambda\mathbf{X})] = \sum_{i=0}^{n} \exp(\lambda i)\binom{n}{i} p^i(1-p)^{n-i} = (1 - p + p\exp(\lambda))^n$
- $P\{\mathbf{X} > n/2\} \le \exp(-n\lambda/2)(1 - p + p\exp(\lambda))^n$ for *all* values of $\lambda > 0$
- Minimum value of RHS is $(2\sqrt{p(1-p)})^n$ at $\lambda = \ln((1-p)/p)$
- For $p = 10^{-2}$, n should be at least 8 to ensure that $P\{\mathbf{X} > n/2\} < 10^{-5}$.
- **Chebyshev's Inequality:** $\mathbf{X}$ is a random variable with finite mean μ and finite variance σ^2. Then, $P\{|\mathbf{X}-\mu| \ge \alpha\} \le \frac{\sigma^2}{\alpha^2}$. Equivalently, $P\{|\mathbf{X}-\mu| \ge k\sigma\} \le \frac{1}{k^2}$

- **Chebyshev's Inequality:** $\mathbf{X}$ is a random variable with finite mean μ and finite variance σ^2. Then, $P\{|\mathbf{X}-\mu| \geq \alpha\} \leq \frac{\sigma^2}{\alpha^2}$. Equivalently, $P\{|\mathbf{X}-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$



- $P\{|\mathbf{X}-\mu| \geq \alpha\}$ = shaded area = $\int_{-\infty}^{\infty} g(u)f(u)du \leq \int_{-\infty}^{\infty} h(u)f(u)du$
- $P\{|\mathbf{X}-\mu| \geq \alpha\} \leq \int_{-\infty}^{\infty} h(u)f(u)du$ where $h(u) = (u-\mu)^2/\alpha^2$
- $P\{|\mathbf{X}-\mu| \geq \alpha\} \leq \frac{\sigma^2}{\alpha^2}$
- Chebyshev's inequality is applicable in general but provides a weak bound
- One-sided Chebyshev inequalities are sometimes useful:
- $P\{\mathbf{X} \geq \mu + \alpha\} \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}$
- $P\{\mathbf{X} \leq \mu - \alpha\} \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}$
- **The Weak Law of Large Numbers**
  $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ denote n i.i.d. (independent identically distributed) random variables with mean μ and variance σ^2.
- Let $\mathbf{Z} = \frac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n}{n}$ = average of the n values
- $E[\mathbf{Z}] = \frac{1}{n}\sum_{i=1}^{n} E[\mathbf{X}_i] = \mu$
- $\mathrm{var}(\mathbf{Z}) = \frac{1}{n^2}\mathrm{var}(\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(\mathbf{X}_i) = \frac{\sigma^2}{n}$
- The Weak Law of Large Numbers is an application of Chebyshev's inequality to $\mathbf{Z}$
- **Weak Law of Large Numbers:** For any $\alpha > 0$,

$$P\{|\mathbf{Z} - \mu| > \alpha\} = P\left\{\left|\frac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n}{n} - \mu\right| > \alpha\right\} \leq \frac{\sigma^2}{n\alpha^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

- **Example:** Let A = event of probability p, i.e. P(A) = p. The experiment is repeated n times
  $\mathbf{X}_i$ = indicator function of A on the i-th trial = $\begin{cases} 1, & \text{if A occurred on i-th trial,} \\ 0, & \text{if } A^c \text{ occurred on i-th trial.} \end{cases}$
- $E[\mathbf{X}_i] = p$; $\quad$ $\text{var}(\mathbf{X}_i) = p(1-p)$
- $\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n$ = number of times A occurred on n trials
- $\mathbf{Z} = \dfrac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n}{n}$ = relative frequency of A
- $E[\mathbf{Z}] = p$; $\quad$ $\text{var}(\mathbf{Z}) = p(1-p)/n$
- **Weak Law of Large Numbers:**
  P{relative frequency differs from probability p by more than α} $\le \dfrac{p(1-p)}{n\alpha^2} \to 0$ as $n \to \infty$
- If α is small, we need to choose a large n
- p(1–p) has maximum value 0.25 at p = 1/2
- In statistical applications, the interval $[p-\alpha, p+\alpha]$ is called a confidence interval; we are confident that $\mathbf{Z}$ is in the confidence interval
- $1 - \dfrac{p(1-p)}{n\alpha^2}$ or $1 - \dfrac{0.25}{n\alpha^2}$ is called the confidence level, or confidence coefficient
- $P\{|\mathbf{Z} - p| \le \alpha\} = P\{p - \alpha \le Z \le p + \alpha\} = P\{Z \text{ in confidence interval}\} \ge 1 - \dfrac{p(1-p)}{n\alpha^2}$
  = confidence level
- Confidence levels of 95% and 99% are commonly used
- If we want high accuracy in estimating p, then we need to use a small α
- To be confident in our "highly accurate" estimate, we must use a large n (expensive)
- Else, we must accept a lower confidence level or less accuracy
- There is a confidence trick going on here
- We repeat an experiment 1000 times and observe the relative frequency of A to be 0.483
- What is P(A) = p?
- We can't give a definitive answer, but will give a range of values for p
- Probability theory says that if $\alpha = 0.1$, then $0.25/1000(0.1)^2 = 0.025 = 2.5\%$ of all such measurements of relative frequencies will be outside the range $p \pm 0.1$
- We refuse to repeat the 1000 trials many times
- We have a relative frequency measurement of 0.483
- We say that we are 97.5% confident that p is in the range 0.483 ± 0.1
- It is *not guaranteed* that p is in this range
- Our particular set of 1000 trials could well be among the "bad" 2.5%

- **The Strong Law of Large Numbers:**

  $\mathbf{X}_i$ = indicator function of A on the i-th trial = $\begin{cases} 1, & \text{A occurred on i-th trial,} \\ 0, & \text{A}^c \text{ occurred on i-th trial.} \end{cases}$
- We repeat the experiment indefinitely often and observe the values of the $\mathbf{X}_i$
- The particular sequence that we are observing could have one of three properties:
- The relative frequency of A converges to p
- The relative frequency of A converges to some $p^* \neq p$
- The relative frequency of A does not converge at all
- **Example:**
  One 1 followed by ten 0's followed by a hundred 1's followed by a thousand 0's, …
- The relative frequency will oscillate back and forth between nearly 1 and nearly 0
- Strong Law of Large Numbers says this has 0 probability
- P{relative frequency converges to $p^* \neq p$ or relative frequency does not converge at all} = 0
- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ denote i.i.d. random variables with mean μ and variance σ^2. Then,

  for any $\alpha > 0$ $P\left\{\lim_{n \to \infty} \frac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n}{n} - \mu > \alpha\right\} = 0$
- The *Weak Law of Large Numbers* asserts that the limit of a probability is 0,

  viz. $\lim_{n \to \infty} P(\text{something}) = 0$
- The *Strong Law of Large Numbers* asserts that the probability of a limit is 0,

  viz. $P(\lim_{n \to \infty} \text{something}) = 0$
- The Central Limit Theorem states (loosely speaking) that the cumulative probability distribution function (CDF) of the sum of a large number of independent random variables is approximately Gaussian.
- **The Central Limit Theorem:**

  $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n, \ldots$ denote i.i.d. random variables with mean μ and variance σ^2. Then,

  for all α, $P\left\{\frac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n - n\mu}{\sigma\sqrt{n}} \leq \alpha\right\} \to \Phi(\alpha)$ as $n \to \infty$
- $\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n$ has mean $n\mu$ and variance $n\sigma^2$
- $\frac{\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n - n\mu}{\sigma\sqrt{n}}$ is a random variable with mean 0 and variance 1
- Its CDF converges to unit Gaussian CDF as n increases
- Convergence of the CDF is much better near the central lobe than in the tails of the distribution
- One should not estimate tail probabilities such as

  $P\{\mathbf{X} > \mu + 3\sigma\}$ via the central limit theorem, but estimating $P\{|\mathbf{X}-\mu| \leq 3\sigma\}$ is OK
- Estimating a probability of 0.99999 as 0.99995 leads to a very small relative error
- Estimating the tail probability 0.00001 as 0.00005 might be disastrous!
- One should not estimate tail probabilities via central limit theorem
- The CDF of $\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n$ is not converging to anything
- However, in practical use, we just treat $\mathbf{X}_1+\mathbf{X}_2+ \ldots +\mathbf{X}_n$ as an $N(n\mu, n\sigma^2)$ random variable
- One should not estimate tail probabilities via central limit theorem