10. Binomial Example
   a. Standardize to standard normal: \( X_i \sim \text{Bin}(1, \pi) \), 
      \( Y_i = (X_i - \pi) / \sqrt{\pi(1 - \pi)} \)
   b. \[ P(Y_i = y) = \begin{cases} 
   -\pi / \sqrt{\pi(1 - \pi)} & \text{with probability } 1 - \pi \\
   1 - \pi / \sqrt{\pi(1 - \pi)} & \text{with probability } \pi 
   \end{cases} \]
   c. Here \( E(Y_i^2) = (1 - 2\pi)(\pi(1 - \pi)^{-1/2} \rightarrow \pi \)
      i. Because 
      \[ E(Y_i^2) = (-\pi^3(1 - \pi) + (1 - \pi)^3\pi(\pi(1 - \pi))^{-3/2} \]
      \[ = \pi(1 - \pi)(-\pi^2 + (1 - \pi)^2)(\pi(1 - \pi))^{-3/2} \]
      \[ = (1 - 2\pi)(\pi(1 - \pi))^{-1/2} \]
   d. \( E(Y_i^3) = 0 \) when \( \pi = 1/2 \), and goes to infinity as 
      \( \pi \rightarrow 0 \) or \( \pi \rightarrow 1 \).

WMS: 7.5

11. A Further Refinement to the Central Limit Theorem
   a. Represent discrete distribution as a bar plot
      i. Bar centered at data point.
      ii. Bars abut halfway between the points.
      iii. Height makes probability on point equal to area of 
          bar.
      iv. Approximate sum of probabilities by the appropriate 
          area under a normal.
   b. Adjust area under normal curve to match area of 
      probability to be calculated.
      i. Adjust ends of interval to midpoint between support 
          points. See Fig. 47.
      ii. \( P(X \geq x) = P(X \geq x - \Delta/2) \).

b. Let \( \sigma \) represent standard deviation of \( Y_j \).
   c. Effect on argument to \( \Phi \) is \( \Delta/(2\sqrt{n}\sigma) \).
   d. Goes to zero as \( n \rightarrow \infty \)
   e. In respect to power of \( n \), effect size is comparable to 
      that of skewness.

B. Laws of Large Numbers
1. Probability Bounds on the Distance between the Average 
   and Expectation
   a. Suppose that
      i. \( Y_i \) are iid random variables
      ii. Expectation \( \mu \)
      iii. Variance \( \sigma^2 < \infty \) (we will drop this)
   b. Then \( E(\bar{Y}_n) = \mu \), \( V(\bar{Y}_n) = \sigma^2/n \)
   c. Take \( \epsilon > 0 \).
   d. Tchebysheff’s Inequality:
      \[ P(|\bar{Y}_n - \mu| > \epsilon) \leq \sigma^2/(n\epsilon^2) \]
   e. Central limit argument: \( P(|\bar{Y}_n - \mu| > \epsilon) = \)
      \[ P \left( \frac{|\bar{Y}_n - \mu|}{\sigma / \sqrt{n}} > \sqrt{n}\epsilon / \sigma \right) \approx 2\Phi(-\sqrt{n}\epsilon / \sigma) \]
   f. CLT bound is smaller than Tchebysheff bound
      i. but it is only approximate and not certain.
      ii. Can be made certain if there exists a finite third 
          moment. See Fig. 48.

2. Statement of Weak Law of Large Numbers
   a. Select
      i. \( Y_i \) iid with finite expectation \( \mu \)
      ii. \( \epsilon > 0 \)
      iii. \( \bar{Y}_n = \sum_{i=1}^{n} Y_i/n \) as before.
   \[ P(X \leq x) = P(X \leq x + \Delta/2), \text{ for } \Delta \text{ the } \]
   difference between support points.
   • Often \( \Delta = 1 \).
   iii. Correction to range is called continuity correction.
   c. Example: \( Y_m \sim \text{Bin}(m, \pi), m = 20, \pi = .4 \).
      i. \( P(X \leq 8) = P(X \leq 8.5) = \frac{5680072674457}{3536743140625} = .595599 \)
      ii. \( \Phi((8 - 20 \times .4)/\sqrt{20 \times .4 \times .6}) = .5 \)
      iii. \( \Phi((8.5 - 20 \times .4)/\sqrt{20 \times .4 \times .6}) = .593 \).

12. Continuity correction is less important for large samples
   a. Let \( \Delta \) represent the space between values of \( Y_j \).

Fig. 48: Bounds on Probability of 
Difference from Expectation Exceeding \( \epsilon \)

---

Lecture 25 225 Lecture 25 226

---

Fig. 47: Approximate Probability Calculation for Binomial
C. Waiting Times

1. Probability Generating Functions

a. Let \( Y \) be a random variable taking values on the non-negative integers.

i. Suppose \( Y_j \sim T_2 \) are independent.

ii. \( E(|Y_j|) = \lim_{y \to \infty} 2 \int_0^y \frac{Ct}{(1 + t^2/2)} dt = 4C \lim_{y \to \infty} (1 - 1/\sqrt{2 + x^2/2}) < \infty \)

iii. \( E(|Y|^2) = \lim_{y \to \infty} 2 \int_0^y C t^2 (1 + t^2/2)^{-3/2} dt \geq 2C \int_\infty^\infty \sqrt{t} dt = \infty \).

b. Let \( q_Y(t) = E(t^Y) = \sum_{y=0}^\infty p_Y(y)t^y \).

ii. Finite at least for \( t \in [0, 1] \).

c. Probabilities of \( Y \) can be determined from derivatives of \( q \)

i. \( \frac{d^n}{dt^n} q(t) = \sum_{y=0}^\infty [y(y-1) \cdots (y-k+1) p_Y(y)] t^{y-k} \).

ii. Because one of these factors is zero if \( y < k \), \( \frac{d^n}{dt^n} q(t) = \sum_{y=k}^\infty [y(y-1) \cdots (y-k+1) p_Y(y)] t^{y-k} \).

iii. \( \frac{d^n}{dt^n} q(t) \bigg|_{t=0} = (k-1) \cdots 1 \times p_Y(k) = k! p_Y(k) \).

2. Probability Generating Functions for Sums

a. If \( Z = X + Y \), \( X \perp Y \) and \( Y \) take integer values, then \( q_Z(t) = E(t^Z) = E(t^X)E(t^Y) = q_X(t)q_Y(t) \).

i. Hence probability generating functions can be used to calculate probabilities for sums of random variables.

3. A Generating Function for Tail Probabilities

a. Let \( q_j = \sum_{k=j}^\infty p_Y(k) \) be the probabilities \( P(Y > j) \)

i. Distinction \( > \) is important because the distribution of \( Y \) is discrete.

ii. \( q_0 = 1 - P(Y = 0) \).

b. Let \( Q_Y(t) = \sum_{j=0}^\infty q_j t^j \)

c. Then \( Q_Y(t) = (1 - q_Y(t))/(1 - t) \) for \( t \neq 1 \)

ii. If \( E(Y) < \infty \), then \( \lim_{y \to \infty} y(1 - F_Y(y)) = 0 \), since distribution must have tails smaller than Cauchy.

\( \text{c. Then } E(Y) = \int_0^\infty y(1 - F_Y(y)) dy \).


a. Player makes a series of iid bets with payoffs \( X_i \)

i. \( p_X(x) = \begin{cases} +1 & \text{with probability } \rho \\ -1 & \text{with probability } 1 - \pi \end{cases} \).

ii. Cumulative winnings are \( S_n = \sum_{i=1}^n X_i \)

iii. Player stops as soon as \( S_n = 1 \).

iv. Let \( N \) be the random variable representing this stopping time.

v. \( P(N = 0) = 0 \).

vi. \( P(N = 1) = \pi \).

b. If \( N > 1 \), the first bet was a loss, and the first time being up by one will happen by first returning to zero, and then going up by one after returning to zero.

i. Let \( L = \min \{t \mid \sum_{j=1}^t X_j = 1\} \)

\( M = \min \{m \mid \sum_{j=L+1}^m X_j = 1\} \).

ii. Then for \( m > 0 \), \( P(N = m + 1 | N > 1) = P(L + M = m) \).

iii. Then \( \sum_{m=1}^\infty P(N = m + 1 | N > 1) t^m = \sum_{m=1}^\infty P(L + M = m) t^m \).

iv. Then \( \sum_{m=1}^\infty P(N = m + 1) / P(N > 1) t^m = q_M(t) q_L(t) \).

v. Then \( \sum_{m=1}^\infty P(N = m + 1) t^{m+1} / P(N > 1) t = q_M(t) q_L(t) \).
vi. Then \((1 - \pi)^{-1}t^{-1}(\sum_{m=0}^{\infty} P(N = m)t^m - \pi t) = q_M(t)q_L(t)\)

vii. Then \((1 - \pi)^{-1}t^{-1}(q_N(t) - \pi t) = q_N(t)q_N(t)\)

viii. Then \(q_N(t) = \frac{1 - \sqrt{1 - 4\pi(1 - \pi)t^2}}{2(1 - \pi)t} .\)

- Taylor expansion gives 
  \(\pi t + \left(\pi^2 - \pi^3\right)t^3 + 2\left(\pi^5 - 2\pi^4 + \pi^3\right)t^5 + O(t^7) .\)
- It’s impossible for first gain of 1 to happen on an even number.
- We can get as many of these probabilities as we need this way.