Math 191 Notes, 2003 September 30

30 minute quiz on Thursday, October 9 (not Oct 7 as stated in the syllabus), covers Chapter 1 and Chapter 2.

Independence

Pairwise independence isn't good enough.

Random Variable

A guinea pig is a rodent not a pig. A random variable is a function not a variable. To be precise,

$$XX : \Omega \to \mathbb{R}$$

There is no mention of probability. In order for this to be useful, we need certain events to be able to be assigned probabilities. The events of interest are: $\{\omega \in \Omega | XX(\omega) \leq x\}$.

For example, flip a coin, then another coin, giving the space $\{HH, HT, TH, TT\}$. Let XX_1 be: the number of heads, then $\{2, 1, 1, 0\}$. Let XX_2 be: start with \$1 bet everything on heads twice in a row, final assets $\{4, 0, 0, 0\}$. Let XX_3 be: bet \$1 on first flip, \$2 on second flip – net gain $\{3, -1, 1, -3\}$. Let XX_4 be: bet \$2, then \$4 on 2nd if heads first, \$1 on 2nd if tails first – net gain $\{6, -2, -1, -3\}$. Let YY be: pick a random real in $[0, 1]$ and square it $\mathbb{P}(\frac{1}{4}) = 0$. That's why we need to use intervals, not equals, because otherwise the actual probability will be zero.

Distribution Function

A distribution function introduces probability.

$$F : \mathbb{R} \to [0, 1]$$

$$F(x) = \mathbb{P}(XX \leq x)$$

Case X_2: see diagram, we must assign values for every real number. Case X_1: see diagram.

Take any number in $[0, 1]$, square it. Probability distribution of having anything less than the result is the following graph.

Any more complicated, and we need to use integrals.

One last case: Bernoulli. This is the case where $\mathbb{P}(H) = p, \mathbb{P}(T) = 1 - p, XX(H) = 1, XX(T) = 0$. Indicator function expressing event A, indicator function:

$$I_A(\omega) = \{1 \text{ if } \omega \in A, 0 \text{ if } \omega \in A^c\}$$

For example, take 3 coin flips. Make 2x2 tables.

Law of Averages, Weak Law of Large Numbers

$$S_3 = I_1 + I_2 + I_3$$

Assume $p = 1/2$ Think about S_n/n for very large n. We would like to see the graph for that, so that it is likely that the vertical jump should be mostly at p. We should like to prove that

$$P(p - \epsilon \leq \frac{S_n}{n} \leq p + \epsilon)$$

We expect that $\lim_{n\to\infty}$ of the above will be 1.
Proof! We would like to calculate the following probability, where $q = 1 - p$, and m is the smallest integer $\geq n(p + \epsilon)$. The strategy is to extend the lower limit of summation to 0 so that you can use the binomial theorem, and the extra factors that we are allowing are factors $\alpha^k \beta^{n-k}$:

$$\mathbb{P}\left(\frac{S_n}{n} \geq p + \epsilon\right) = \sum_{k=m}^{n} \binom{n}{k} p^k q^{n-k}$$

Choose $\lambda > 0$, multiply each term by $e^{\lambda(k-n(p+\epsilon))} = e^{-\lambda n\epsilon} e^{-\lambda np} e^{\lambda kp} e^{\lambda kq}$.

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} \geq p + \epsilon\right) &\leq e^{-\lambda n\epsilon} \sum_{k=m}^{n} \binom{n}{k} p^k e^{\lambda kq} q^{n-k} e^{-\lambda(n-k)p} \\
&\leq e^{-\lambda n\epsilon} \sum_{k=0}^{n} \binom{n}{k} \left(pe^{\lambda q}\right)^k \left(qe^{-\lambda p}\right)^{n-k} \\
&= e^{-\lambda n\epsilon} \left(pe^{\lambda q} + qe^{-\lambda p}\right)^n
\end{aligned}$$

Aside: $e^x \leq e^{x^2} + x$ for all real x.

$$\begin{aligned}
f(x) &= e^{x^2} + x - e^x \\
f'(x) &= 2xe^{x^2} + 1 - e^x \\
f''(x) &= 4x^2 e^{x^2} + 2e^{x^2} - e^x
\end{aligned}$$

The proof of the fact that this is greater than zero is that $(x - 1/2)^2 \geq 0$, then exponentiate both sides after moving stuff around and use the fact that $e^{1/4} < 2$.
Back to the proof.

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} \geq p + \epsilon\right) &= e^{-\lambda n\epsilon} \left(pe^{\lambda q} + qe^{-\lambda p}\right)^n \\
&\leq e^{-\lambda n\epsilon} \left(pe^{\lambda^2 q^2} + p\lambda q + qe^{-\lambda^2 p^2} - q\lambda p\right)^n
\end{aligned}$$

$$\leq \quad e^{-\lambda n\epsilon}\left(pe^{\lambda^2}+qe^{\lambda^2}\right)^n = e^{-\lambda n\epsilon}e^{\lambda^2 n}$$

Minimize the exponent by choosing $\lambda = \epsilon/2$, then we have:

$$\mathbb{P}\left(\frac{S_n}{n} \geq p+\epsilon\right) \quad \leq \quad e^{-n\epsilon^2/4}$$

This proof was by Berstein.