

Problem Session 3

1. *Expectation from cdf.* For a positive random variable X imagine how its cdf look like.
- a. Using the representation $X = F^{-1}(U)$ where U is uniform on $(0, 1)$ show that $E(X)$ is the area above the curve of cdf and below the $y = 1$ line:

$$E(X) = \int_0^\infty (1 - F(x))dx = \int_0^\infty P(X > x)dx \quad (1)$$

- b. What is the discrete analogue of this?
- c. Use this to find the mean and variance of the exponential and geometric distributions.
- d. How can you extend this to non-positive random variables?
2. *Coupon collecting.* Each box of chocolate by a certain brand comes with a different gift card. There are a total of n different cards and you want to collect them all. Each box is equally likely to contain any one of the cards. What is the expected number of boxes you must buy in order to obtain the complete set of cards.
3. *Coupon collecting continued.* Let T_n be the number of boxes to get a complete set.
- a. Find a formula for standard deviation of T_n .
- b. Show that the SD is less than a constant times n .
- c. Use Chebychev to show that T_n will differ from $n \log(n)$ by only a small multiple of n .
4. *Conditional expectation.* Let Y be the number of heads in X tosses of a fair coin, where X is a fair die roll.
- a. Find the conditional expectation of Y given $X = x$.
- b. Find $E(Y)$.
- c. Find $E(X|Y = 2)$.
5. *Conditional expectation continued.* S_n be the number of successes in n trials, what is $E(S_m|S_n = k)$ for $m \leq n$, what is $E(S_m|S_n)$? If X and Y are independent what is $E(X + Y|X = x)$.
6. *Bounds.* On a list of million numbers the average is 10, and the average of squares of those numbers is 101.
- a. Find an upper bound on the number of entries that are 14 or more.
- b. What if you know that the distribution of those numbers is symmetric about 10?
- c. What if the distribution is approximately normal?