3.2: Prob. Distributions for Discrete RVs

Taylor

University of Virginia

We'll work with a general discrete rv in this section. We need to do this before we start using specific discrete rvs (i.e. Binomial, Poisson, etc.)

Definitions

A **probability distribution** is a function that shows how the total probability of 1 is distributed among the possible values of a rv (say X).

A **probability mass function** is a probability distribution for a discrete random variable. For each possible number x that X can take on, it gives you $P(X = x)$.

Note

Note: we talked about probability function on $\mathcal{S}$ before. How does that relate to this probability function on the range of X? Well, for a discrete rv: $P(X = x) = P(\{s \in \mathcal{S} : X(s) = x\})$

Note: All those rules we talked about in chapter 2 still apply. Check as many as you want to.

Note: knowing the distribution allows you to do essentially anything you want to (i.e. find the probability of any event, compute moments, etc)

Note: PMFs are usually displayed in one of two ways: a table, or a line graph or a probability histogram

Definitions

When we're dealing with random variables and we want to specify a distribution, typically it isn't feasible to just pick numbers for the probability of each outcome...what if there are an infinite amount? Fortunately there are many probability distributions where we only have to specify one or two **parameters**, and then it fills out the rest.

A **parameter** (for a discrete rv) is a quantity that can change (certain) distribution (probability mass) functions $P(X = x) = p(x)$. Let's say there is only one parameter in our situation, and call it α. To make this dependence on α explicit, sometimes we write our pmf as $p_\alpha(x)$ or $p(x; \alpha)$.

Example

Example 3.1 on page 104: we observe the gender of newborns in a hospital until a boy (B) is born. So $\mathcal{S} = \{B, GB, GGB, GGGB, \ldots\}$. There are a countably inifinite number of possible situations. What if we just assume that there is one probability that an individual birth results in a boy, and all births are independent from one another? Let's call $P(B) = p$.

$X =$ number of births observed. Then $p(x) = (1 - p)^{x-1}p$, $x > 0$

If our assumptions are true, we just have to pick a good value for p

The **cumulative distribution function (cdf)**:
$F_X(x) = P(X \leq x) = \sum_{k \leq x} p(k)$

Example

In our last example, we had $p(x) = (1-p)^{x-1}p$. The associated cdf is

$$
\begin{aligned}
F_X(x) &= \sum_{k=1}^{x}(1-p)^{k-1}p \\
&= p\sum_{k=0}^{x-1}(1-p)^k \\
&= p\left[\frac{1-(1-p)^x}{1-(1-p)}\right] \\
&= p\left[\frac{1-(1-p)^x}{p}\right] \\
&= 1-(1-p)^x
\end{aligned}
$$

We're using that formula for the partial sum of a geometric series here..

Example

Instead of going from pmf to cdf, we can also go from cdf to pmf...

$$P(X = k) = F_X(k) - F_X(k - 1)$$

$$
\begin{aligned}
p(3) &= P(X = 3) \\
&= P\left[(X \leq 3) \cap (X \leq 2)'\right] \\
&= P(X \leq 3) + P((X \leq 2)') - P\left[(X \leq 3) \cup (X \leq 2)'\right] \\
&= P(X \leq 3) + P((X \leq 2)') - P\left[(X \leq 3) \cup (X > 2)\right] \\
&= F_X(3) + (1 - F_X(2)) - 1 \\
&= F_X(3) - F_X(2)
\end{aligned}
$$

Proposition

Here's a more general result:

For any a and b, with $a \leq b$

$$P(a \leq X \leq b) = F_X(b) - F_X(a-)$$

Note: $F_X(x-) = \lim_{\epsilon \to 0} F_X(x - \epsilon)$

Note: A lot of times $F_X(x-) = F_X(x-1)$