Introduction to Probability
Lecture 4: More discrete distributions – Poisson, Geometric, Negative Binomial, Hypergeometric
Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
e-mail: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk
Outline

Poisson discrete random variable

Geometric discrete random variable

Negative binomial discrete random variable

Hypergeometric discrete random variable
Preliminaries:

The natural exponent $e$

$e$ is a mathematical constant AKA the Euler number. $e$ is very important for exponential functions. Here are some important identities:

\[
e \approx 2.71828
\]
\[
e = \sum_{n=0}^{\infty} \frac{1}{n!}
\]
\[
e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n
\]
\[
e^{-\lambda} = \lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n
\]
\[
e^r = \lim_{n \to \infty} \left(1 + \frac{r}{n}\right)^n
\]
\[
e^r = \sum_{n=0}^{\infty} \frac{r^n}{n!}
\]
Binomial RV example: large \( n \), small \( p \)

We are trying to predict footfall in a store. We know, based on previous data, that on average 8 people enter the store per hour. What is the probability of \( k \) people entering the store in the next 1 hour?

1. Break an hour into \textbf{minutes}.
   - At each \textit{minute}, independent Bernoulli trial with 1 for a person entering the store and 0 for nobody entering the store.
   - \( X \) is a Binomial RV: # people entering in an hour, so \( \mathbb{E}[X] = np = \lambda = 8 \).
   - \( X \sim \text{Bin}(n = 60, p = \lambda/n) \), \( \mathbb{E}[X] = np = \lambda = 8 \).
   - \( X \sim \text{Bin}(n = 60, p = \lambda/n) \), so \( P[X=k] = \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \)

2. Break an hour into \textbf{milliseconds}.
   - At each \textit{millisecond}, independent Bernoulli trial: 1 for enter, 0 for not enter.
   - \( X \) is a Binomial RV: # people entering in an hour, so \( \mathbb{E}[X] = np = \lambda = 8 \).
   - \( X \sim \text{Bin}(n = 3600000, p = \lambda/n) \), so \( P[X=k] = \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \)

3. Break an hour into \textbf{infinitely small units}.
   - At each \textit{unit}, independent Bernoulli trial: 1 for enter, 0 for not enter.
   - \( X \) is a Binomial RV: # people entering in an hour, so \( \mathbb{E}[X] = np = \lambda = 8 \).
   - \( X \) is a Binomial RV: # people entering in an hour, thus \( P[X=k] = \lim_{n \to \infty} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} \)

What if 2 people enter in the same minute?

What if 2 people enter in the same millisecond?
Computing Binomial in the limit

\[
P[X = k] = \lim_{n \to \infty} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k}
\]

\[
= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \lambda^k \left( 1 - \frac{\lambda}{n} \right)^n
= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} \lambda^k \left( \frac{1}{1 - \frac{\lambda}{n}} \right)^k
\]

\[
= \lim_{n \to \infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{e^{-\lambda}}{(1 - \frac{\lambda}{n})^k}
\]

as \( n \to \infty \)

\[
\frac{n(n-1)\cdots(n-k+1)}{k!} \approx \frac{n^k}{k!} = 1
\]

Therefore, in our store footfall example: the probability of \( k \) people entering the store in the next 1 hour is:

\[
P[X = k] = \lim_{n \to \infty} \binom{n}{k} \left( \frac{\lambda}{n} \right)^k \left( 1 - \frac{\lambda}{n} \right)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}
\]

\[
P[X = k] = (1) \frac{\lambda^k}{k!} e^{-\lambda}
\]
A Poisson RV $X$ approximates Binomial where $n$ is large, $p$ is small, and $\lambda = np$ is "moderate". Thus we no longer need to know $n$ and $p$, we only need to provide rate $\lambda$. $X$ is the number of successes over the duration of the experiment.

\[ X \sim \text{Pois}(\lambda) \]

Range: \{0, 1, 2, \ldots\}

PMF:
\[ P[ X = k ] = \frac{\lambda^k}{k!} e^{-\lambda} \]

Expectation:
\[ E[ X ] = \lambda \]

Variance:
\[ V[ X ] = \lambda \]

Examples: # earthquakes in a given year, # goals scored during a 90 minute football game, # misprints per page in a book, # emails per day.

Key idea: Divide time into a large number of small increments. Assume that during each increment, there is some small probability of the event happening (independent of other increments).
Earthquake example

Example

Suppose there are an average of 2.79 major earthquakes in the world each year. What is the probability of getting 3 major earthquakes next year?

Define RVs: \( \lambda = 2.79, k = 3, X \sim \text{Pois}(2.79) \)
Poisson paradigm

- Poisson approximates Binomial when $n$ is large, $p$ is small, and $\lambda = np$ is "moderate".
- Different interpretations of "moderate". Commonly accepted ranges are:
  - $n > 20$ and $p < 0.05$
  - $n > 100$ and $p < 0.1$
- Poisson is Binomial in the limit: $\lambda = np$ where $n \to \infty$, $p \to 0$. 

![Histogram of Poisson and Binomial distributions]

- **Poisson** discrete random variable
Poisson expectation

\[ \text{PMF: } k \in \{0, 1, 2, \ldots \infty\}; \ P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \]

\[ E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \]

\[ = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \text{ (let } i = k - 1) \]

\[ = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \lambda e^{-\lambda} e^{\lambda} = \lambda \]
Poisson variance

\[
E\left[ X^2 \right] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \quad \text{(let } i = k - 1) \\
= \lambda \sum_{i=0}^{\infty} (i + 1) \frac{\lambda^i}{i!} e^{-\lambda} = \lambda \left( \sum_{i=0}^{\infty} \frac{i \lambda^i}{i!} e^{-\lambda} + \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \right) = \lambda(\lambda + 1) \quad \text{thus}
\]

\[
V[ X ] = E[ X^2 ] - (E[ X ])^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda
\]

\[
E\left[ X^k \right] = \lambda E\left[ (X + 1)^{k-1} \right]
\]
Bernoulli, Poisson, and random processes

- A Poisson process is a model for a series of discrete events where the average time between events is known, but the exact timing of events is random.
  - The arrival of an event is independent of the event before (waiting time between events is memoryless).
  - The average rate (events per time period) is constant.
  - Two events cannot occur at the same time: each sub-interval of a Poisson process is a Bernoulli trial that is either a success or a failure.

- Example: your website goes down on average twice per 60 days; calling a help centre; movements of stock price...
Outline

Poisson discrete random variable

Geometric discrete random variable

Negative binomial discrete random variable

Hypergeometric discrete random variable
Geometric discrete random variable

$X$ is a geometric RV if $X$ is a number of independent Bernoulli trials until the first success and $p$ is the probability of success on each Bernoulli trial.

$X \sim \text{Geo}(p)$

Range: \{1, 2, \ldots\}

PMF: $P[X = n] = (1 - p)^{n-1}p$

Expectation: $E[X] = \frac{1}{p}$

Variance: $V[X] = \frac{1 - p}{p^2}$

Examples: tossing a coin ($P[\text{head}] = p$) until first heads appears, generating bits with $P[\text{bit} = 1] = p$ until first 1 is generated.
Geometric RV properties

PMF ($E_i$ is the event that the $i$-th trial succeeds):

$$
P[ X = n ] = P[ E_1^c E_2^c \ldots E_{n-1}^c E_n ] =$$
$$= P[ E_1^c ] P[ E_2^c ] \ldots P[ E_{n-1}^c ] P[ E_n ] =$$
$$= (1 - p)^{n-1} p$$

CDF ($P[ X > n ]$ is the probability that at least the first $n$ trials fail):

$$
P[ X \leq n ] = 1 - P[ X > n ] =$$
$$= 1 - P[ E_1^c E_2^c \ldots E_n^c ] =$$
$$= 1 - P[ E_1^c ] P[ E_2^c ] \ldots P[ E_n^c ] =$$
$$= 1 - (1 - p)^n$$
You roll a fair 6-sided die until it comes up with # 6. What is the probability that it will take 3 rolls?

Let $X$ be a RV for # of rolls. Probability for any # on die is $\frac{1}{6}$.

Define RVs: $X \sim Geo\left(\frac{1}{6}\right)$, want $P[ X = 3 ]$.

Solve:
Outline

Poisson discrete random variable

Geometric discrete random variable

Negative binomial discrete random variable

Hypergeometric discrete random variable
Negative binomial discrete random variable

$X$ is a negative binomial RV if $X$ is the number of independent Bernoulli trials until $r$ successes and $p$ is the probability of success on each trial.

$$X \sim \text{NegBin}(r, p)$$

Range: $\{r, r + 1, \ldots\}$

PMF: $P[X = n] = \binom{n - 1}{r - 1}(1 - p)^{n-r} p^r$

Expectation: $E[X] = \frac{r}{p}$

Variance: $V[X] = \frac{r(1 - p)}{p^2}$

Examples: tossing a coin until $r$-th heads appears, generating bits until the first $r$ 1’s are generated.

Note: $\text{Geo}(p) = \text{NegBin}(1, p)$. 
A PhD student is expected to publish 2 papers to graduate. A conference accepts each paper randomly and independently with probability $p = 0.25$. On average, how many papers will the student need to submit to a conference in order to graduate?

Answer

---

Example (not real life!)
Let $X \sim \text{NegBin}(m, p)$ and $Y \sim \text{NegBin}(n, p)$ be two independent RVs. Define a new RV as $Z = X + Y$. Find PMF of $Z$.

Answer

- Need to show that $Z \sim \text{NegBin}(m + n, p)$.

- Consider the sequence of independent events tossing a coin with $\mathbb{P}[\text{heads}] = p$.

- Let $X$ be a RV for # of coin tosses until $m$ heads are observed. Thus $X \sim \text{NegBin}(m, p)$.

- Now, continue to toss a coin after $m$ heads are observed, until $n$ more heads are observed. Thus, for this part of the sequence, $Y \sim \text{NegBin}(n, p)$.

- Looking at it from the beginning we tossed independently the coin until we observed $m + n$ heads, thus $Z = X + Y$ and thus $Z \sim \text{NegBin}(m + n, p)$.

- Note: if $X_1, X_2, \ldots, X_m$ are $m$ independent $\text{Geo}(p)$ RVs, the the RV $X = X_1 + X_2 + \cdots + X_m$ has $\text{NegBin}(m, p)$ distribution.
Outline

Poisson discrete random variable

Geometric discrete random variable

Negative binomial discrete random variable

Hypergeometric discrete random variable
Hypergeometric discrete random variable

$X$ is a hypergeometric RV that samples $n$ objects, without replacement, with $i$ successes (random draw for which the object drawn has a specified feature), from a finite population of size $N$ that contains exactly $m$ objects with that feature.

$$X \sim \text{Hyp}(N, n, m)$$

Range: $\{0, 1, \ldots, n\}$

PMF: 
$$P[ X = i ] = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}$$

Expectation: 
$$E[ X ] = n \frac{m}{N}$$

Variance: 
$$V[ X ] = n \frac{m}{N} \left( 1 - \frac{m}{N} \right) \left( 1 - \frac{n-1}{N-1} \right)$$

Example: an urn has $N$ balls of which $m$ are white and $N - m$ are black; we take a random sample without replacement of size $n$ and measure $X$: # of white balls in the sample.
Survey sampling

Example

A street has 40 houses of which 5 houses are inhabited by families with an income below the poverty line. In a survey, 7 houses are sampled at random from this street. What is the probability that: (a) none of the 5 families with income below poverty line are sampled? (b) 4 of them are sampled? (c) no more than 2 are sampled? (d) at least 3 are sampled?

Let $X$: # of families sampled which are below the poverty line.

$X \sim Hyp(N = 40, n = 7, m = 5)$. 
## Summary of discrete RV

<table>
<thead>
<tr>
<th></th>
<th>$Ber(p)$</th>
<th>$Bin(n, p)$</th>
<th>$Pois(\lambda)$</th>
<th>$Geo(p)$</th>
<th>$NegBin(r, p)$</th>
<th>$Hyp(N, n, m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMF</td>
<td>$P[X = 1] = \frac{p}{p}$</td>
<td>$P[X = k] = \binom{n}{k}p^k(1-p)^{n-k}$</td>
<td>$P[X = k] = \frac{\lambda^k}{k!}e^{-\lambda}$</td>
<td>$P[X = n] = \frac{(1-p)^{n-1}p}{n}$</td>
<td>$P[X = n] = \frac{(1-p)^{n-1}p^r}{r}$</td>
<td>$P[X = i] = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$</td>
</tr>
<tr>
<td>$E[X]$</td>
<td>$p$</td>
<td>$np$</td>
<td>$\lambda$</td>
<td>$\frac{1}{p}$</td>
<td>$\frac{r}{p}$</td>
<td>$n \frac{m}{N}$</td>
</tr>
<tr>
<td>$V[X]$</td>
<td>$p(1-p)$</td>
<td>$np(1-p)$</td>
<td>$\lambda$</td>
<td>$\frac{1-p}{p^2}$</td>
<td>$\frac{r(1-p)}{p^2}$</td>
<td>$n \frac{m}{N} \left(1 - \frac{m}{N} \right) \left(1 - \frac{n-1}{N-1} \right)$</td>
</tr>
<tr>
<td>Descr.</td>
<td>1 experiment with prob $p$ of success</td>
<td>$n$ independent trials with prob $p$ of success</td>
<td># successes over experiment duration, $\lambda = np$ rate of success</td>
<td># independent trials until first success</td>
<td># independent trials until $r$ successes</td>
<td># successes of drawing item with a feature (without replacement) in a sample of size $n$ from a population of size $N$ with $m$ items with the feature</td>
</tr>
</tbody>
</table>