

Special probability distributions

Lecture 11: part 1

STAT GR5203

Overview

- Many naturally occurring phenomena follow special, known probability distributions
 - Examples: binomial, geometric
- These distributions are often related to one another
- In this lecture
 - Where do these known distributions arise
 - Disciplines where these distributions are used/applied
 - Relationships between groups of distributions

Bernoulli, Binomial, and Multinomial distributions

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

- How it arises: an experiment, or trial, whose outcome can be classified as either a success or failure is performed
- $X = \mathbb{I}\{\text{outcome is a "success"}\}$
- p.m.f.: $p(x) = p^x(1 - p)^{1-x}$ for $x \in \{0,1\}$
 - Parameters: $p \in [0,1]$ (probability of success)
- Examples:
 - Genetics: the probability of inheriting specific genetic traits (Punnett square)
 - Neuroscience: within a very short time window, a neuron fires with some probability p or remains dormant
 - Physics: within a very short time window, a radioactive atom decays with some probability p or does not decay

Bernoulli, Binomial, and Multinomial distributions

$$X \sim \text{Binom}(n, p)$$

- How it arises: n independent Bernoulli trials are performed. Each of these trials has probability p of success
- X = number of successes in n trials
- p.m.f.: $p(x) = \binom{n}{x} p^x (1 - p)^{1-x}$ for $x \in \{0, 1, \dots, n\}$
 - Parameters: $n \in \{1, 2, 3, \dots\}$ (number of trials), $p \in [0, 1]$ (probability of success)
- Examples following a $\text{Binom}(n, p)$ distribution
 - Election polling: $100p$ % the voting populace supports candidate A. X = the number of supporters of A among a sample of n voters follows a $\text{Binom}(n, p)$ distribution
 - Clinical trials: A drug has a $100p$ % efficacy rate on a specific subpopulation. X = the number of patients among a sample of n individuals from this subpopulation for whom the drug works

Bernoulli, Binomial, and Multinomial distributions

$$(X_1, X_2, \dots, X_K) \sim \text{Multinomial}(n, p_1, p_2, \dots, p_K)$$

- How it arises: n independent trials are performed, where the trial has K possible outcomes. Each of these trials has probability p_1 of outcome 1, p_2 of outcome 2, ..., p_K of outcome K
- $(X_1, X_2, \dots, X_K)$ = (Number occurrences of outcome 1, Number occurrences of outcome 2, ..., Number occurrences of outcome K) in n trials
- p.m.f.: $p(x_1, x_2, \dots, x_K) = \binom{n}{x_1, x_2, \dots, x_K} p_1^{x_1} p_2^{x_2} \dots p_K^{x_K}$ for $(x_1, \dots, x_K) \in \{0, 1, \dots, n\}^K$ such that $\sum_{j=1}^K x_j = n$
 - Parameters: $n \in \{1, 2, 3, \dots\}$ (number of trials), $(p_1, \dots, p_K)$ (outcome probabilities such that $p_j \in [0, 1]$ for all $j \in \{1, \dots, K\}$ and $\sum_{j=1}^K p_j = 1$)
- Example: topic modeling
 - Given a set of documents, we want to identify some common topics among them
 - The model **assumes** that the distribution of words for each topic follows a multinomial distribution
 - Example: if the topic is "pets" and the set of words is {dog, cat, car, bus,...} then we would expect p_{dog} and p_{cat} to be higher than p_{car} or p_{bus}

Bernoulli, Binomial, and Multinomial distributions

Relationships

- A $\text{Binom}(1, p)$ r.v. is a $\text{Bern}(p)$ r.v.
- **Exercise:** Suppose that $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$, where X and Y are independent. Show that $X + Y \sim \text{Bin}(n + m, p)$
- If $(X_1, X_2, \dots, X_K) \sim \text{Multinomial}(n, p_1, p_2, \dots, p_K)$, then $X_i \sim \text{Binom}(n, p_i)$

Hypergeometric distribution

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

- How it arises: Given a population of size N with two types of objects, m being of type 1 and $N - m$ of type 2, we sample of n extracted objects.
- X = number of objects of type 1 in a sample of n
- p.m.f.: $p(x) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$ for $\max\{0, n - (N - m)\} \leq x \leq \min\{n, m\}$
 - Parameters: $n \in \{0, 1, 2, \dots\}$ (number of sampled objects), $N \in \{0, 1, 2, \dots\}$ (number of total objects), $m \in \{0, 1, 2, \dots\}$ (number of objects of type 1)
- Example: poker hands, balls from urns

Example: Hypergeometric and Binomial distributions

Sampling with/without replacement

- Manufacturing: a factory produced N light bulbs, $m \leq N$ of which are known to be defective. The manufacturer chooses $n \leq N$ bulbs to inspect
- Define a "success" as finding a defective bulb. Let X = the number of defectives found in the sample
 - If the bulbs are chosen **with replacement**, then $X \sim \text{Binom}\left(n, \frac{m}{N}\right)$
 - If the bulbs are chosen **without replacement**, then $X \sim \text{Hypergeom}(n, N, m)$
- **Exercise:** Suppose that $X \sim \text{Hypergeom}(n, N, m)$. Let $m = np$ for some $p \in [0, 1]$. Show that as $N \rightarrow \infty$,
$$\mathbb{P}(X = x) \rightarrow \binom{n}{x} p^x (1 - p)^{n-x}$$
- Intuition, if we have more total objects, sampling without replacement more closely resembles sampling with replacement

Poisson and Exponential distributions

$$X \sim \text{Pois}(\lambda)$$

- How it arises: consider a process where (i) events occur at a known mean rate; (ii) the occurrence of events in one time interval is independent of the occurrence of events in another (disjoint) time interval; and (iii) events cannot occur simultaneously (Poisson process)
- X = the number of occurrences in a fixed interval of time*
- p.m.f.: $p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ for $x \in \{0, 1, 2, \dots\}$
 - Parameters: $\lambda > 0$ (average number of occurrences in a fixed interval of time)
- Examples:
 - Physics: within a very short time window, the number of atoms in a radioactive sample that decay in one hour follows a Poisson distribution
 - Insurance pricing: in a given month, the number of car accidents in a region follows a Poisson distribution. This information can be used to set insurance premiums

*the unit need not be time, it can be space, for example

Limit: Binomial $\rightarrow$ Poisson

- Notice that, like the Binomial distribution examples, the Poisson distribution examples consider the occurrence/non-occurrence of events
- The Poisson distribution also arises as a limit of the Binomial distribution
 - Break up a time interval into very small subintervals such that at most one event can occur in a subinterval
 - The occurrence of an event in each subinterval follows a Bernoulli distribution
 - The number of events in the larger interval is the sum of occurrences in the subintervals, which are independent and Bernoulli (sum of independent Bernoullis is a Binomial)
- Binomial-Poisson Limit Theorem: Suppose $X_n \sim \text{Binom}(n, np_n)$, where $p_n \in [0, 1]$ for all n and $np_n \rightarrow \lambda$. Then as $n \rightarrow \infty$, $X_n \approx \text{Pois}(\lambda)$
 - **Exercise:** show that $\mathbb{P}(X_n = x) \rightarrow \frac{e^{-\lambda} \lambda^x}{x!}$ as $n \rightarrow \infty$

Poisson and Exponential distributions

$$X \sim \mathbf{Exp}(\lambda)$$

- How it arises: "distance" between events following a Poisson process (e.g., time between events)
- X = distance between events
- p.d.f.: $f(x) = \lambda e^{-\lambda x}$ for $x > 0$
 - Parameters: $\lambda > 0$ ("rate")
- Examples:
 - Physics: time between atom decays
 - Insurance pricing: time between car accidents

Geometric and Negative Binomial distributions

- How they arise: independent Bernoulli trials are repeatedly performed. Each of these trials has probability p of success
- Geometric distribution
 - X = number of failures before first success
 - p.m.f.: $p(x) = (1 - p)^{x-1}p$ for $x \in \{1, 2, \dots\}$
 - Parameters: p (probability of success)
- Negative binomial distribution
 - X = number of failures before r^{th} success
 - p.m.f.: $p(x) = \binom{r+x-1}{x} p^r (1 - p)^x$ for $x \in \{0, 1, \dots\}$
 - Parameters: $p \in [0, 1]$ (probability of success), $r \in \{1, 2, 3, \dots\}$ (number of successes to observe)
- Example (sports analytics): a basketball player has a $100p$ % chance of making a free throw
 - The number of shots until getting a free throw follows a **geometric** distribution
 - The number of shots until getting a certain number of free throws follows a **negative Binomial** distribution

Memoryless: Geometric and Exponential distributions

- Recall from HW 2: you showed that if $T \sim \text{Exp}(\lambda)$, then $\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t)$ ("Memoryless property")
 - Example: suppose a chemical reaction has the potential to occur at any time. We observe no reaction at time s . This does not change the probability of observing a reaction after waiting t more minutes.
 - Interpretation: T is the waiting time since time s for an event to occur. Whatever did not happen up to s does not affect what will happen in t more time
- The geometric distribution is a *discrete* distribution that is also memoryless
 - Example: I flip a coin 10 times and observe no heads. The probability of needing 3 more tosses to see a head is the same as needing at least 3 tosses to see a head without having considered the previous 10 tosses
- Idea in both cases: "past failures do not dictate future success/failures." Very inspirational!

Gamma and Exponential distributions

Relationship between them

- If $X \sim \text{Exp}(\lambda)$, then we also have that $X \sim \text{Gamma}(1, \lambda)$
- If $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(\lambda)$, then $\sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$
- If $X_i \sim \text{Gamma}(t_i, \lambda)$ and the $X_1, \dots, X_n$ are mutually independent, then $\sum_{i=1}^n X_i \sim \text{Gamma}(\sum_{i=1}^n t_i, \lambda)$
- In general, if $X \sim \text{Gamma}(a, \lambda)$, then $f(x) = \frac{\lambda^a e^{-\lambda x} x^{a-1}}{\Gamma(a)}$ for $x \geq 0$
 - Parameters: $\alpha > 0$ (shape parameter), $\lambda > 0$ (rate parameter)

Normal, χ^2 , and Gamma distributions

- Next part of the lecture: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$
- If $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(0,1)$, then $X_1^2 + \dots + X_n^2 \sim \chi_n^2$ ("Chi-squared" distribution)
 - The Chi-squared distribution will come up in statistical inference/hypothesis testing. The parameter, n , is called the "degrees of freedom"
 - The χ_n^2 distribution is equivalent to the $\text{Gamma}(n/2, 1/2)$ distribution

Uniform and Beta distributions

- Recall the uniform distribution: if $X \sim \text{Unif}(a, b)$, then $f(x) = \frac{1}{b-a} \mathbb{I}\{a < x < b\}$
- Recall the Beta distribution: If $Y \sim \text{Beta}(a, b)$, then $f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}$ for $y \in [0, 1]$, and $a, b > 0$
 - This distribution will come up in the context of Bayesian statistics
- Consider $(X_1, \dots, X_n) \stackrel{iid}{\sim} \text{Unif}(0, 1)$. Let us reorder the sample into $(X_{(1)}, \dots, X_{(n)})$ so that $X_{(k)}$ is the k^{th} largest value
 - $(X_{(1)}, \dots, X_{(n)})$ are called order statistics (will come up in inference)
 - **Exercise:** show that $X_{(k)} \sim \text{Beta}(k, n - k + 1)$