Parametric families of (univariate) distributions

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Overview

There are many types of data for which we would like to have a suitable distributional model:

- Discrete: categorical, ordinal, counts
- Continuous: durations, generic errors

In practice, one usually works with a suitable **parametric family** of densities, and we do the same here.[1]

[1] In Advanced Statistics III, we will discuss **nonparametric** approaches.

Parametric models

- We will use the generic notation $f(x;\theta)$ (and $F(x;\theta)$ for the cdf):
  - This denotes a family of densities for random variable X.
  - A given value for θ pins down a specific member of the family.
- The admissible values θ of the parameters are called the **parameter space** and will be denoted by Ω. (Vector $\boldsymbol{\theta}$ allowed for as well.)
- Each family is more suitable for certain tasks and comes with specific parameter interpretations/symbols.[2] So use with care.

[2] In Advanced Statistics II, we shall discuss **estimation** of θ given sample data.

Today's outline

Parametric families of (univariate) distributions

1. Models for categorical data
2. Models for counts
3. Models for durations
4. Up next

Outline

The discrete uniform distribution

Family Name: Discrete Uniform	
Parameterization	$N \in \Omega = \{N : N \text{ is a positive integer}\}$
Density Definition	$f(x; N) = \frac{1}{N}\mathbb{I}_{\{1,2,\ldots,N\}}(x)$
Moments	$\mu = (N+1)/2, \sigma^2 = \left(N^2 - 1\right)/12, \mu_3 = 0$
MGF	$M_X(t) = \sum_{j=1}^{N} e^{jt}/N$

This is (only) suitable if your outcomes are equally likely.

Example

Consider the experiment of rolling a die. The pdf of the number of dots facing up is $f(x; N = 6) = \frac{1}{6}\mathbb{I}_{\{1,2,\ldots,6\}}(x)$, and belongs to the family of discrete uniforms.

The Bernoulli distribution

Family Name: Bernoulli

Parameterization	$p \in \Omega = \{p : 0 \leq p \leq 1\}$
Density Definition	$f(x; p) = p^x (1-p)^{1-x} \mathbb{I}_{\{0,1\}}(x)$
Moments	$\mu = p, \sigma^2 = p(1-p), \mu_3 = 2p^3 - 3p^2 + p$
MGF	$M_X(t) = pe^t + (1-p)$

This works perfectly if you only have two possible outcomes – not necessarily equally likely.

Usually 0 stands for one out of two categories and 1 for the other.

The binomial distribution

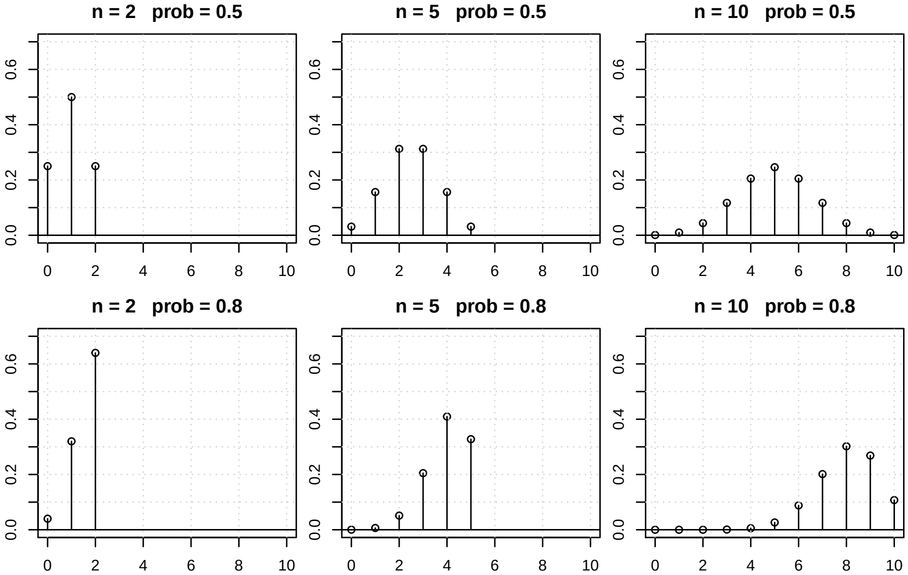
Family Name: Binomial

Parameterization	$(n,p) \in \Omega = \{(n,p) : n$ is a positive integer, $0 \leq p \leq 1\}$
Density Definition	$f(x;n,p) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, & \text{for } x = 0,1,2,\ldots,n \\ 0 & \text{otherwise} \end{cases}$
Moments	$\mu = np, \sigma^2 = np(1-p), \mu_3 = np(1-p)(1-2p)$
MGF	$M_X(t) = (1-p+pe^t)^n$

The binomial density is used to model an experiment that consists of n independent repetitions of a Bernoulli-type experiment with a success probability p.

The quantity of interest x is the total number of successes in n of such Bernoulli trials. (Compare the MGFs)

Some specific binomial pmfs (or discrete pdfs)



Example

What is the probability of obtaining at least one '6' in four rolls of a fair die?

- This experiment can be modeled as a sequence of $n = 4\ \mathrm{Ber}(p)$ trials with success probability $p = 1/6 = \mathrm{P}(\text{6 dots face up})$.
- Define the random variable $X =$ total number of 6s in four rolls.
- Then $X \sim \mathrm{Binom}(n = 4\ , p = 1/6)$ and

$$\begin{aligned}\mathrm{P}(\text{at least one `6'}) &= \mathrm{P}(X > 0) = 1 - \mathrm{P}(X = 0)\\ &= 1 - \binom{4}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^4 = .518.\end{aligned}$$

The multinomial distribution

Family Name: Multinomial

Parameterization	$(n, p_1, \ldots, p_m) \in \Omega = \{(n, p_1, \ldots p_m) : n$ is a positive integer, $0 \leq p_i \leq 1, \forall\, i, \sum_{i=1}^m p_i = 1\}$
Density Definiton	$f(x_1, \ldots, x_m; n, p_1, \ldots, p_m) = \begin{cases} \frac{n!}{\prod_{i=1}^m x_i!} \prod_{i=1}^m p_i^{x_i} & \text{for } x_i = 0, 1, 2, \ldots, n\, \forall\, i, \quad \sum_{i=1}^m x_i = n \\ 0 & \text{otherwise} \end{cases}$
Moments	$\mu_i = np_i, \sigma_i^2 = np_i(1-p_i), \mu_{3,i} = np_i(1-p_i)(1-2p_i)$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$
MGF	$M_X(t) = \left(\sum_{i=1}^m p_i e^{t_i}\right)^n$

The quantities of interest $x_1, ..., x_m$ are the total numbers of each of the m different possible outcomes in n independent repetitions of the experiment.

Note that the range of the random vector $(X_1, ..., X_n)$ is given by $\mathrm{R}(\boldsymbol{X}) = \{(x_1, ..., x_n) : x_i \in \{0, 1, ..., n\} \forall i, \quad \sum_{i=1}^m x_i = n\}$.

Outline

The negative binomial distribution

Family Name: Negative Binomial (Pascal)

Parameterization	$(r,p) \in \Omega \{(r,p) : r$ is a positive integer, $0 < p < 1\}$
Density Definition	$f(x;r,p) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{for } x = r, r+1, r+2, \ldots \\ 0 & \text{otherwise} \end{cases}$
Moments	$\mu = \frac{r}{p}, \sigma^2 = \frac{r}{p^2}(1-p), \mu_3 = \frac{r}{p^3}\left((1-p) + (1-p)^2\right)$
MGF	$M_X(t) = e^{rt} p^r \left(1 - (1-p)e^t\right)^{-r}$ for $t < -\ln(1-p)$

This models (randomly many) independent $\mathrm{Ber}(p)$ experiments/trials.

- The quantity of interest x is the number of Bernoulli trials which are necessary to obtain r successes.
- Compared to the binomial, the number of trials and the number of successes are reversed w.r.t. what is random and what is a parameter.

A special case: the geometric distribution

Set $r = 1$ such that

$$f(x;p) = p\,(1-p)^{x-1} \quad \text{for} \;\; x = 1, 2, \ldots.$$

The quantity of interest x is the Bernoulli trial at which the first success occurs.[3]

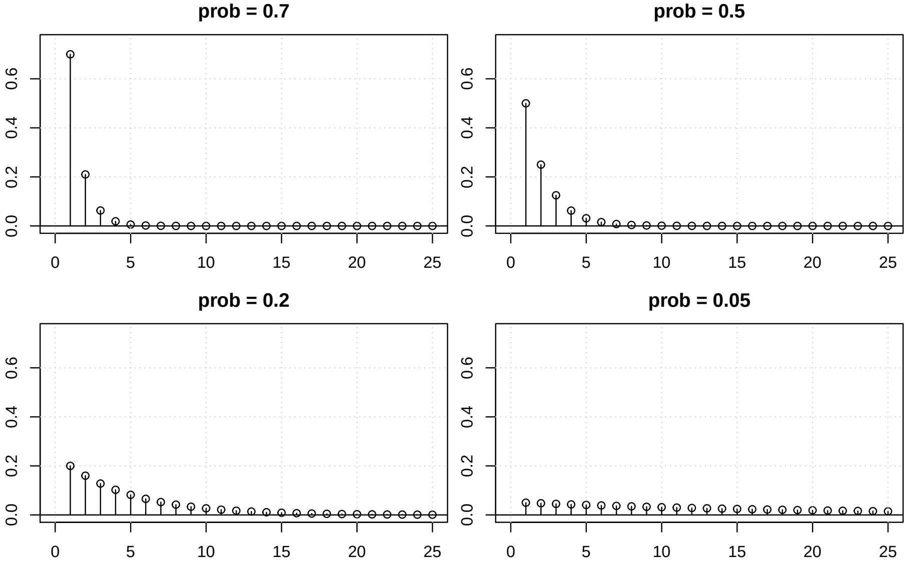
The **geometric distribution** has a property known as the **memoryless property**. It means that for some positive integers i and j we obtain

$$\mathrm{P}(X > i + j | X > i) = \mathrm{P}(X > j).$$

The memoryless property can be interpreted as a *lack-of-aging* property.

[3] Some (e.g. in **R**) take x to be the number of failures before the first success.

Some specific geometric pmfs



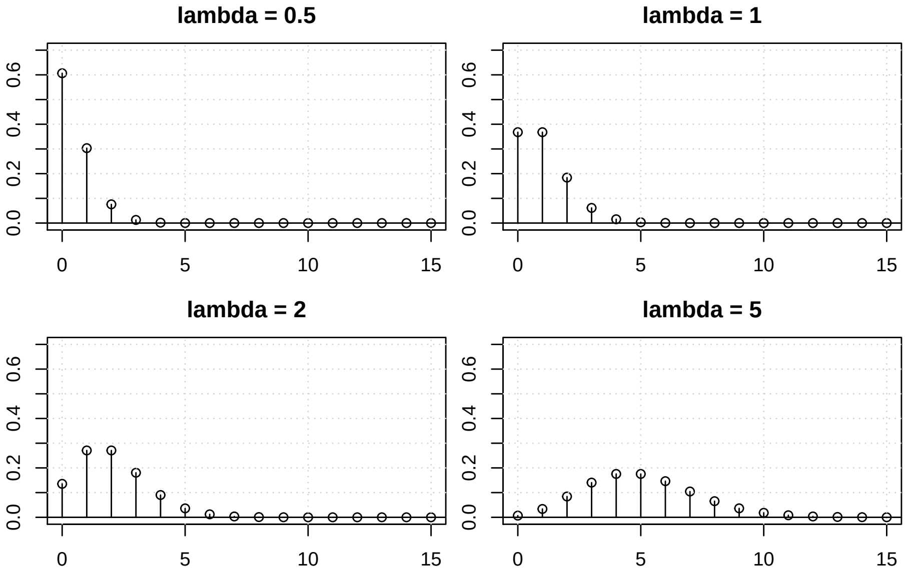
And the last one

Family Name: Poisson

Parameterization	$\lambda \in \Omega = \{\lambda : \lambda > 0\}$
Density Definition	$f(x;\lambda) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & \text{for } x = 0, 1, 2... \\ 0 & \text{otherwise} \end{cases}$
Moments	$\mu = \lambda, \sigma^2 = \lambda, \mu_3 = \lambda$
MGF	$M_X(t) = e^{\lambda\left(e^t - 1\right)}$

The Poisson distribution is also called the law of rare events (see below why).

Some Poisson pmfs



The Poisson process

The Poisson distribution models experiments whose outcomes are governed by the so-called Poisson process:

Definition (Poisson process)

Let an experiment consist of observing the occurrence of a certain event over a time interval $[0, t]$. The experiment follows a Poisson process if:

1) the probability that the event occurs once over a small time interval Δt is approximately proportional to Δt as[a] $\gamma \cdot (\Delta t) + o(\Delta t)$, where $\gamma > 0$,
2) the probability that the event occurs twice or more often over a small time interval Δt is negligible being of order of magnitude $o(\Delta t)$,
3) the numbers of occurrences of the event that are observed in non-overlapping intervals are independent events.

[a] $o(\Delta t)$ stands for *of smaller order than* Δt and means that the values of $o(\Delta t)$ approach zero at a rate faster than Δt. That is $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

... and the formal connection

Theorem (4.1)

Let X be the number of times a certain event occurs in the interval $[0, t]$. If the experiment underlying X follows a Poisson process, then $X \sim \mathrm{Po}(\lambda)$.

- The parameter γ is interpreted as the mean rate of occurrence of the event per unit of time or the intensity of the Poisson process;
- This follows from the fact that for a Poisson variable $\mathrm{E}(X) = \lambda = \gamma t$ such that $\mathrm{E}(X/t) = \gamma$.

A shortcut to the binomial

The Poisson distribution provides an approximation to the probabilities generated by the binomial distribution.

- In fact, the limit of the binomial density as the number of Bernoulli trials $n \to \infty$ is the Poisson density if $np \to \lambda > 0$.
- For a large number of trials n and thus for a small success probability $p = \lambda/n$, we can use

$$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \approx \frac{(np)^x e^{-np}}{x!}.$$

- The Poisson density is relatively easy to evaluate, whereas, for large n, the calculation of the factorial expressions is not.

Outline

The Gamma distribution

Family Name: Gamma

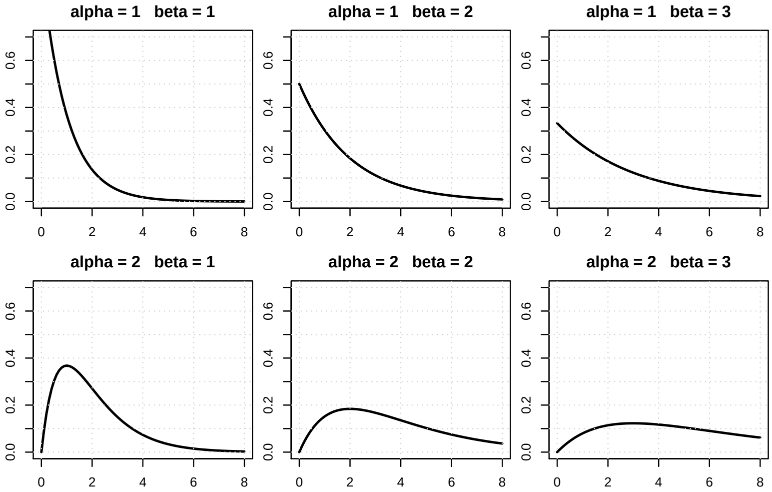
Parameterization	$(\alpha, \beta) \in \Omega = \{(\alpha, \beta) : \alpha > 0, \beta > 0\}$
Density Definition	$f(x; \alpha, \beta) = \frac{1}{(\beta^\alpha \Gamma(\alpha))} x^{\alpha-1} e^{-x/\beta} \mathbb{I}_{(0,\infty)}(x)$, where $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} \mathrm{d}y$.
Moments	$\mu = \alpha\beta, \sigma^2 = \alpha\beta^2, \mu_3 = 2\alpha\beta^3$
MGF	$M_X(t) = (1 - \beta t)^{-\alpha}$ for $t < \beta^{-1}$

The gamma function has the following properties.

- For any real $\alpha > 0$, the gamma function satisfies the recursion $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. This can be verified through integration by parts.
- $\Gamma(1) = \int_0^\infty e^{-y} \mathrm{d}y = 1$ and $\Gamma(1/2) = \pi^{1/2}$.
- If $\alpha > 0$ is an integer, then $\Gamma(\alpha) = (\alpha - 1)!$.

Useful

Since the range of the Gamma distribution is $\mathbb{R}_+$, it's a natural choice for modelling durations. Must accept skewness to the right though.



Some properties

- The Gamma distribution models the waiting time (duration) between occurrences of events under a Poisson process.
- The gamma distribution has an **additivity property**, see below.
- A rescaled gamma distribution is also gamma, see below.

Theorem (4.2)

Let $X_1, ..., X_n$ *be independent RVs with* $X_i \sim \mathrm{Gamma}(\alpha_i, \beta)$*,* $i = 1, ..., n$*. Then* $Y = \sum_{i=1}^n X_i \sim \mathrm{Gamma}(\sum_{i=1}^n \alpha_i, \beta)$*.*

Theorem (4.3)

Let $X \sim \mathrm{Gamma}(\alpha, \beta)$*. Then, for any* $c > 0$*,* $Y = cX \sim \mathrm{Gamma}(\alpha, \beta c)$*.*

(Like before, compare MGFs)

The exponential special case

Gamma Subfamily Name: Exponential

Parameterization	$\theta \in \Omega = \{\theta : \theta > 0\}$
Density Definition	$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta} \mathbb{I}_{(0,\infty)}(x)$
	$f(x;\lambda) = \lambda e^{-\lambda x} \mathbb{I}_{(0,\infty)}(x)$
Moments	$\mu = \theta, \sigma^2 = \theta^2, \mu_3 = 2\theta^3$
MGF	$M_X(t) = (1 - \theta t)^{-1}$ for $t < \theta^{-1}$

A specific application of the exponential distribution is the modeling of the time that passes until a Poisson process produces the first success.

No memory

The exponential distribution has the **memoryless property** (too):

Theorem (4.4)

If $X \sim \mathrm{Exp}(\theta)$*, then* $\mathrm{P}(X > s + t | X > s) = \mathrm{P}(X > t) \; \forall \; (t, s) > 0$.

This indicates that the exponential distribution is not appropriate to model lifetimes for which the failure probability is expected to increase with time.

The χ^2 special case

A further important special case of the gamma distribution, obtained by setting $\alpha = v/2$ and $\beta = 2$, is the **chi-square distribution**.

Gamma Subfamily Name: Chi-Square

Parameterization	$v \in \Omega = \{v : v \text{ is a positive integer}\}$ v is called the **degrees of freedom**
Density Definition	$f(x; v) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2)-1} e^{-x/2} \mathbb{I}_{(0,\infty)}(x)$
Moments	$\mu = v, \sigma^2 = 2v, \mu_3 = 8v$
MGF	$M_X(t) = (1-2t)^{-v/2}$ for $t < \frac{1}{2}$

The chi-square distribution plays an important role in statistical inference.

In particular, (as we will show later) the sum of the squares of v independent standard normal random variables has a χ^2_v-distribution.

Related, though not a duration

Family Name: Beta

Parameterization	$(\alpha, \beta) \in \Omega = \{(\alpha, \beta) : \alpha > 0, \beta > 0\}$
Density Definition	$f(x; \alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1} (1-x)^{\beta-1} \mathbb{I}_{(0,1)}(x)$,
	where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} \, \mathrm{d}x$ is the beta function[4].
Moments	$\mu = \alpha/(\alpha+\beta), \quad \sigma^2 = \alpha\beta / \left[(\alpha+\beta+1)(\alpha+\beta)^2\right]$,
	$\mu_3 = 2(\beta-\alpha)(\alpha\beta) / \left[(\alpha+\beta+2)(\alpha+\beta+1)(\alpha+\beta)^3\right]$
MGF	$M_X(t) = \sum_{r=1}^{\infty} \left(B(r+\alpha, \beta) / B(\alpha, \beta)\right) (t^r/r!)$

The beta density can be used to model experiments whose outcomes are coded as real numbers on the interval $[0, 1]$. It has obvious applications in modeling random variables representing proportions.

[4]Some useful properties of the beta function include the fact that $B(\alpha, \beta) = B(\beta, \alpha)$ and $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$.

A particular case: the uniform distribution

Family Name: Continuous Uniform

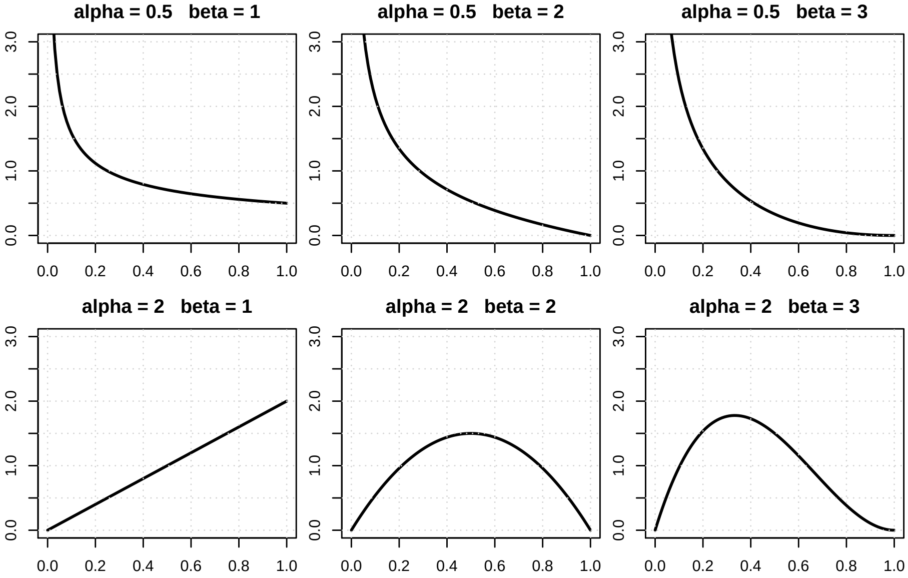
Parameterization	$(a,b) \in \Omega = \{(a,b) : -\infty < a < b < \infty\}$
Density Definition	$f\left(x;a,b\right) = \frac{1}{b-a}\mathbb{I}_{[a,b]}(x)$
Moments	$\mu = \left(a+b\right)/2, \sigma^2 = \left(b-a\right)^2/12, \mu_3 = 0$
MGF	$M_X(t) = \begin{cases} \frac{e^{bt}-e^{at}}{(b-a)t} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0 \end{cases}$

Fun fact: let $X \sim F$ be a continuous RV; then, $F(X) \sim \mathrm{Unif}(0,1)$.
Anyway, the $\mathrm{Beta}(1,1)$ distribution is the same as $\mathrm{Unif}(0,1)$.

Example

Spin a wheel of fortune with radius r. The point X at which the wheel stops is uniformly distributed with $a = 0$ and $b = 2\pi r$.

A lot of flexibility



Outline

Coming up

The normal family of distributions