Continuous distributions

Christopher A. Sims
Princeton University
sims@princeton.edu

September 10, 2020

$S = \mathbb{R}^k$

- We've been considering distributions over finite state spaces. Now we start considering distributions with the state space $S = \mathbb{R}^k$, i.e. k-dimensional Euclidean space.

- The simplest version of the finite S theory has every subset with a well defined probability, $S = 2^S$, and every point in S with a well-defined probability. Then
$$P[A] = \sum_{s \in A} P[s] \,.$$

- With a continuous state space, the corresponding simplest case is one where we have a **probability density function** (pdf) p defined on S so that
$$P[A] = \int_{s \in A} p(s)\, ds \,.$$

Defining properties of probability in $\mathbb{R}^k$

- Once we leave the finite state-space case, we have to add two new conditions to the defining properties of probability.

- We need not only that $\mathcal{F}$ contains the union of any two sets in itself, but also that if $\{A_i\}_{i=1}^{\infty}$ is a countable sequence of sets, each in $\mathcal{F}$, then the union of all the sets in the sequence is also in $\mathcal{F}$. (This makes it a σ-**field**".)

- Also, if the sets in $\{A\}_{i=1}^{\infty}$ are disjoint,

$$P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i].$$

  This is called **countable additivity**.

pdf's and expectations of random variables

- Usually we are studying or using the joint distribution of a vector of random variables, say X, which takes values in $\mathbb{R}^k$, and it is convenient to treat the values of X as defining the state space.

- In this case, we can define density functions as functions of the X vector itself, and define the pdf of a set A of X-values as

$$\int_{x \in A} p(x)\, dx \,.$$

- The expectation of X is

$$E[X] = \int x p(x)\, dx \,.$$

Joint and marginal pdf's

- If we have a pdf for a vector $X = (X_1, X_2)$ with two components, the **marginal pdf** for X_1 defines the distribution of X_1 on a reduced state space defined by the values of the X_1 subvector alone.

- The marginal pdf g for X_1 can be derived from the joint pdf for the whole X vector via

$$g(x_1) = \int p(x_1, x_2)\, dx_2 \,.$$

Conditional pdf's and expectations

- The **conditional pdf** $h(x_2 \mid x_1)$ for $X_2 \mid X_1$ can be calculated from the joint and marginal via

$$h(x_2 \mid x_1) = \frac{p(x_1, x_2)}{\int p(x_1, x_2)\, dx_2} = \frac{p(x_1, x_2)}{g(x_1)}\,.$$

Note that h integrates to one in x_2 by construction.

- The **conditional expectation** of a function f of X_2 can be calculated as

$$E[f(X_2) \mid X_1] = \int f(x_2) h(x_2 \mid x_1)\, dx_2$$

Note that as with our discrete-state case, This is a function of X_1, and thus itself a random variable. Also that by construction the law of iterated expectations holds.

pdf's of functions of random variables

- If a one-dimensional x has pdf $p(x)$ and $y = f(x)$, with f strictly monotonic so that we can invert it to find $x = f^{-1}(y)$. Then it is *not* true that the pdf of y is $p(f^{-1}(y))$.

- One way to remember the correct way to derive the pdf of y is to think of the pdf of x as "$p(x)\,dx$" and that of y as "$q(y)\,dy$".

- Then the correct pdf of y is

$$p(f^{-1}(y))\frac{dx}{dy}dy = p(f^{-1}(y))\left|\frac{1}{f'(f^{-1}(y))}\right|dy = q(y)dy$$

Examples of transformations

- A standard distribution is the $\mathrm{Gamma}(p, a)$ distribution, whose pdf is

$$\frac{a^p x^{p-1} e^{-ax}}{\Gamma(p)}$$

on $(0, \infty)$.

- The mean of a random variable with this distribution (i.e. its expectation) is p/a and its **mode**, the value of x at which its density is largest, is $(p-1)/a$. A plot of it appears slide after next.

- If $X \sim \mathrm{Gamma}(p, a)$, what is the pdf of $y = 1/x$?

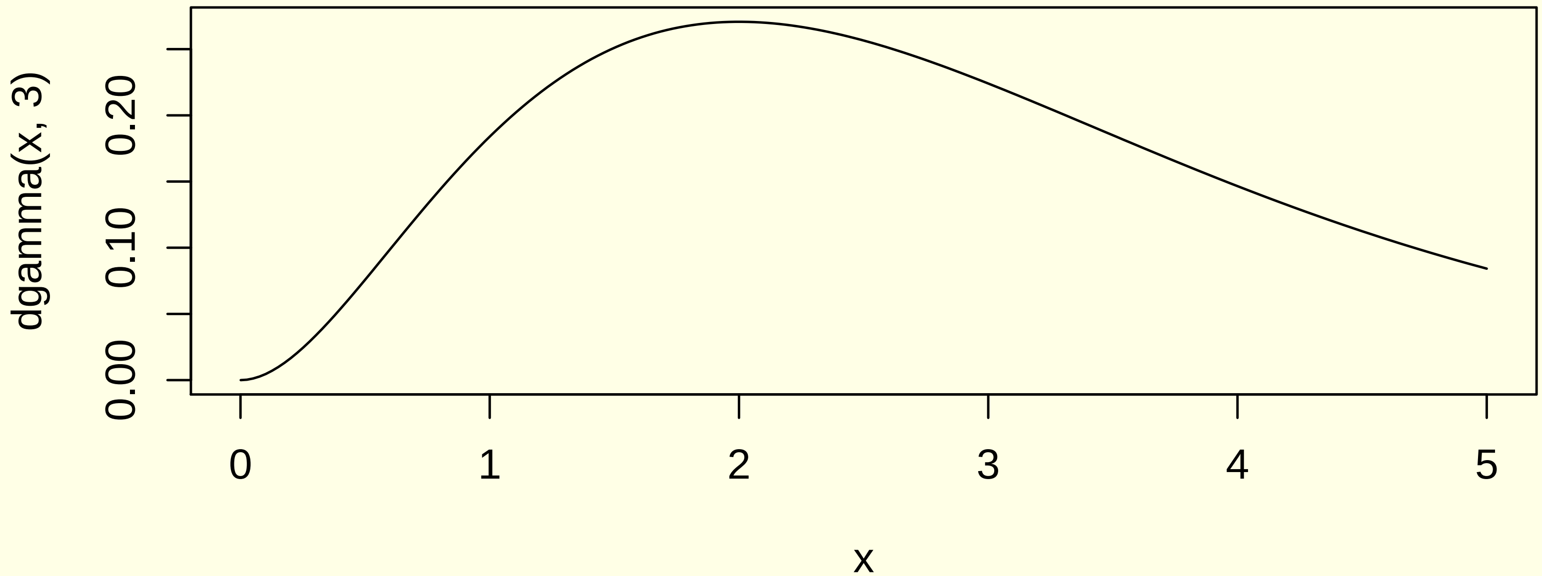
The inverse-gamma, $\mathrm{IG}(p, a)$, distribution

- Note that while x ranges from 0 to ∞, $1/x$ ranges from ∞ to 0, which is why $f'(x)$ is negative in this case.

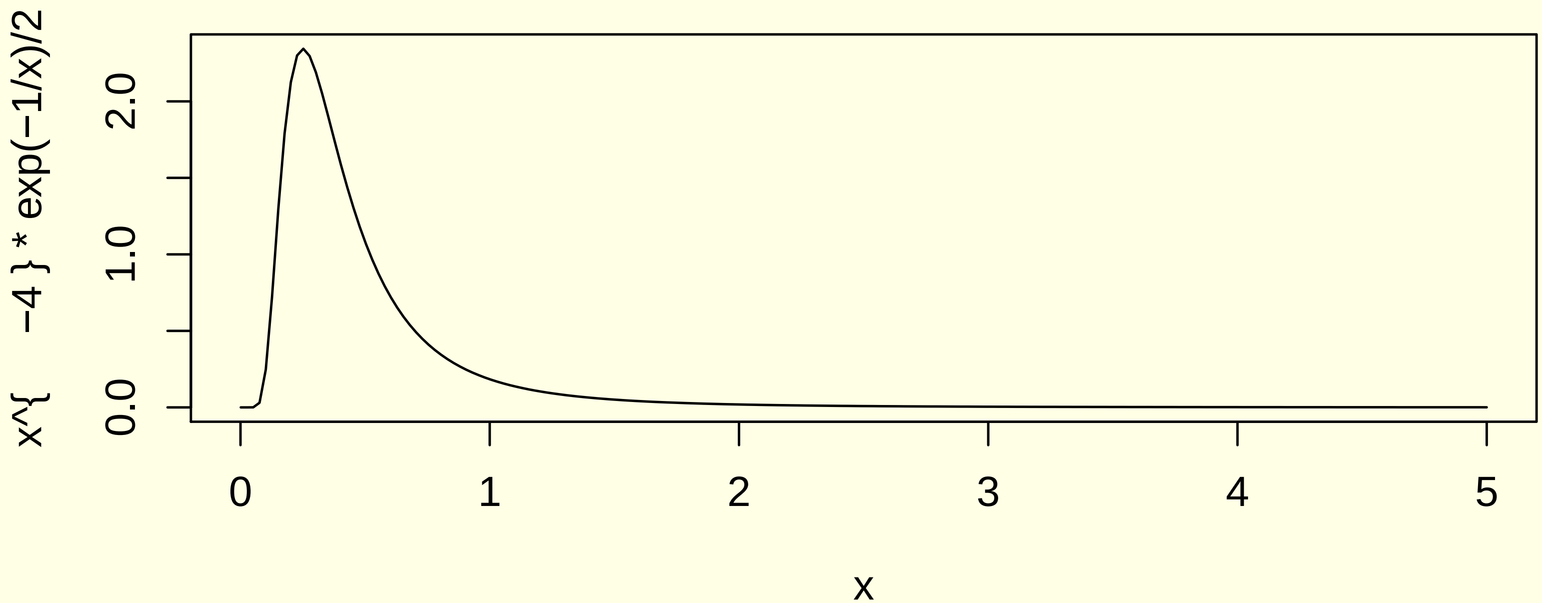
- The density for y, according to the formula two slides back, is

$$\frac{a^p y^{-p+1} e^{-a/y}}{\Gamma(p)} y^{-2}\, dy = \frac{y^{-p-1} e^{-a/y}\, dy}{\Gamma(p)}\,.$$

- This inverse-gamma distribution also crops up frequently in inference. Its mean is $a/(p-1)$ and its mode is at $a/(p+1)$.



dgamma(x, 3)
0.00
0.10
0.20
0
1
2
3
4
5
x



x^{ −4 } * exp(−1/x)/2
0.0
1.0
2.0
0
1
2
3
4
5
x

Multivariate transformations

- Suppose X is a random vector of length k, and $Y = f(X)$ is also a random vector of length k, and $f()$ is one-one, meaning that it has an inverse function $f^{-1}(Y) = X$.

- Then there is an expression for finding the pdf $q()$ of Y from knowledge of $f()$ and the pdf $p()$ of X:

$$q(y)\,dy = p(f^{-1}(y))\left|\frac{\partial f}{\partial x}\right|^{-1} dy\;.$$

Multivariate transforms (cont.)

- This is close to the same form as in the univariate case, except that $\frac{dx}{dy}$ is replaced by the determinant expression, where

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f_i}{\partial x_j}\right] \, .$$

i.e. a matrix in which the i'th row, j'th column is $\partial f_i(x)/\partial x_j$

Classic example of 2d transform

- A distribution we will encounter often is the **joint normal** distribution. It's general k-dimensional form has pdf

$$(2\pi)^{-k/2}\,|\Sigma|^{-1/2}\exp\left(-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right),$$

  where Σ is a positive definite $k \times k$ matrix and μ is a k-dimensional vector. Its mean is μ, which is also its mode. The distribution's name is $N(\mu, \Sigma)$

- Consider the special case where $\Sigma = I$ is 2 by 2 and $\mu = (0, 0)'$. We'll also restrict attention to the upper half plane, so we'll use the pdf restricted to the region where $x_2 > 0$, and thus multiplied by 2 so it integrates to one over that regio

Classic example of 2d transform.2

- A common alternative to the rectangular x, y coordinates to describe points in $\mathbb{R}^2$ is polar coordinates. We index points by their distance ρ from zero and the angle θ that the vector from zero to the point makes with the x axis.

- The transform from rectangular (x_1, x_2) coordinates to (ρ, θ), and its derivative matrix, are

$$\begin{bmatrix}\rho\\ \theta\end{bmatrix} = f(x) = \begin{bmatrix}\sqrt{x_1^2 + x_2^2}\\ \arctan(x_2/x_1)\end{bmatrix} \qquad \frac{\partial f(x)}{\partial x} = \begin{bmatrix}\frac{x_1}{\rho} & \frac{x_2}{\rho}\\ \frac{-x_2}{\rho^2} & \frac{x_1}{\rho^2}\end{bmatrix}$$

Classic example of 2d transform.3

- The term $|\partial f/\partial x|$. whose inverse is called the **Jacobian** term, evaluates to just $1/\rho$. The joint density of ρ and θ is then

$$2 \cdot (2\pi)^{-1} e^{-\frac{1}{2}\rho^2} \rho \, d\rho \, d\theta$$

  on the region where $\rho > 0$ and $\theta \in (0, \pi)$.

- Notice that the pdf of ρ along the line where $\theta = \pi/2$ (indeed along any ray through the origin) is 0 at $\rho = 0$ The rectangular coordinate pdf along the same line is proportional to $e^{-\frac{1}{2}x_2^2}$, which is positive at $x_2 = 0$. Thus the conditional density for x_2 along that line is different from the conditional density for ρ along that line, even though the line is the same set in the upper half plane of $\mathbb{R}^2$ in both cases.

Classic example of 2d transform.4

- This illustrates the point that conditional densities in $\mathbb{R}^k$ are not conditional distributions given *sets*. The upper half of the vertical axis is a set, but it is a set with zero probability under a continuous distribution in $\mathbb{R}^k$.

- One way to gain insight into this example is to realize that $E[\rho \mid \theta = \pi/2]$ is the limit as $\varepsilon \to 0$ of $E[\rho \mid \theta \in (\pi/2 - \varepsilon, \pi/2 + \varepsilon)]$. This latter object conditions on a set, a pie-slice shaped set, which naturally has more weight far from zero than does the thin ribbon $x_1 \in (-\varepsilon, \varepsilon)$ that is the corresponding conditioning set in rectangular coordinates.