Physics 403

Probability Distributions II:
More Properties of PDFs and PMFs

Segev BenZvi

Department of Physics and Astronomy
University of Rochester

Table of Contents

Last Time

- Binomial Distribution
- Poisson Distribution
- Gaussian Distribution
- Central Limit Theorem
- Uniform Distribution
- χ^2 Distribution

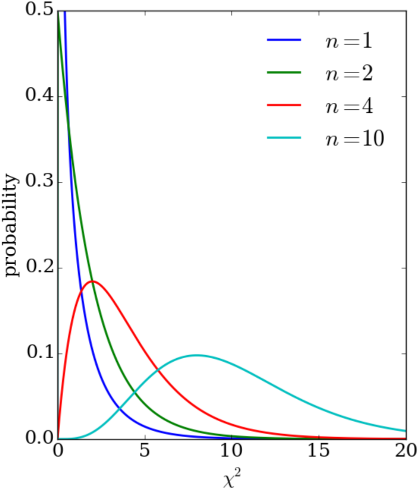
χ^2 Distribution

Using χ^2 to Estimate "Goodness of Fit"

- For n independent Gaussian x_i with means μ_i and variances σ_i^2, the quantity

$$z = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

  follows a χ^2 with n degrees of freedom.
- Notice that z looks like a least-squares estimator for a fit.
- Physicists often use the tail probability of χ^2 as a measure of goodness of fit.

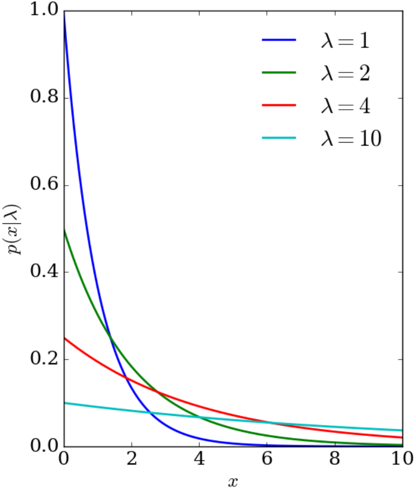


Exponential Distribution

- The exponential PDF is

$$p(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \quad x \geq 0$$

- **Mean**: $\mathrm{E}\,(x) = \lambda$.
- **Variance**: $\mathrm{var}\,(x) = \lambda^2$, RMS: λ
- Lack of memory: $p(t - t_0 | t \geq t_0, \lambda) = p(t|\lambda)$.
- Decay time of unstable particle with lifetime $\lambda \to \tau$
- Lifetime of electrical components, such as lightbulbs



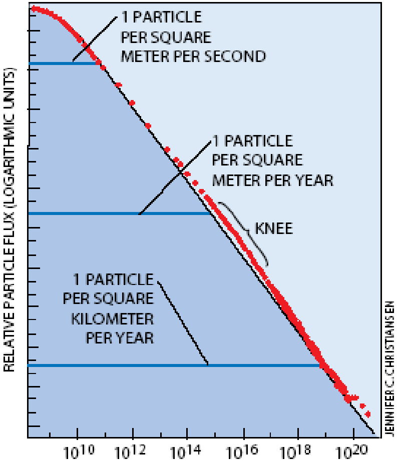
Power Law (Pareto) Distribution

- Power law:

$$p(x|\alpha) = Cx^{-\alpha}$$

- The power law shows up all over physics, and is characteristic of scale invariance, hierarchy, or stochastic generating processes.
- Examples: populations of cities, sizes of lunar impact craters, energies of cosmic rays, sizes of interstellar dust particles, magnitudes of earthquakes, ...

Scientific American, (c) 1998

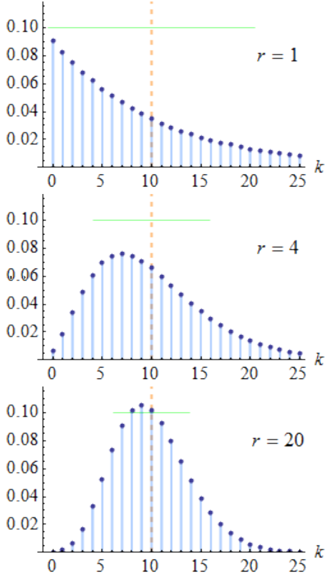


Negative Binomial Distribution

- The negative binomial describes the number of successes in Bernoulli trials up to r failures
- The discrete PDF (actually, PMF) is

$$\binom{k+r-1}{k} p^k (1-p)^r \quad \text{for } k = 0, 1, 2, \ldots$$

- Mean: $pr/(1-p)$
- Variance: $pr/(1-p)^2$
- Sometimes used in place of the Poisson distribution when sample variance > sample mean



Negative Binomial Distribution

Example

Selling cookies (from Wikipedia): a Girl Scout is required to sell boxes of cookies to get a merit badge. There are 30 houses in her neighborhood, and she needs to sell 5 boxes before returning home.

If there is a 40% chance of selling a box at any given house, what is the probability of selling the last box at the n^{th} house?

The neg. binomial describes the probability of k failures and r successes in $k + r$ trials with success on the last trial. Setting $r = 5$, $p = 0.4$, and $n = k + 5$, we can write

$$P(k|r,p) = \binom{k+r-1}{k} p^k (1-p)^r,$$

$$P(n|r=5,p=0.4) = \binom{(n-5)+5-1}{n-5} 0.4^5 0.6^{n-5} = \binom{n-1}{n-5} 2^5 \frac{3^{n-5}}{5^n}$$

Negative Binomial Distribution

Example

What is the probability that the Girl Scout finishes on the 10^{th} house?

$$p(n = 10 | r = 5, p = 0.4) = \binom{9}{5} 2^5 \frac{3^5}{5^{10}}$$
$$\approx 0.1$$

Negative Binomial Distribution

Example

What is the probability that the Girl Scout finishes on or before the 8^{th} house?

She needs to sell 5 boxes, so she must finish at house 5, 6, 7, or 8. Therefore, we sum over these possibilities:

$$\begin{aligned} P(n \leq 8|r,p) &= \sum_{m=5}^{8} P(m|r,p) \\ &= P(5|r,p) + P(6|r,p) + \ldots + P(8|r,p) \\ &\approx 0.010 + 0.031 + 0.055 + 0.077 \\ &\approx 0.173 \end{aligned}$$

Negative Binomial Distribution

Example

What is the probability that the Girl Scout does not sell all her boxes after visiting the whole neighborhood?

We want the probability that she does not finish on houses 5 through 30. The probability that she **does** finish by the last house is

$$P(n \leq 30|r,p) = \sum_{m=5}^{30} P(m|r,p) \approx 0.998.$$

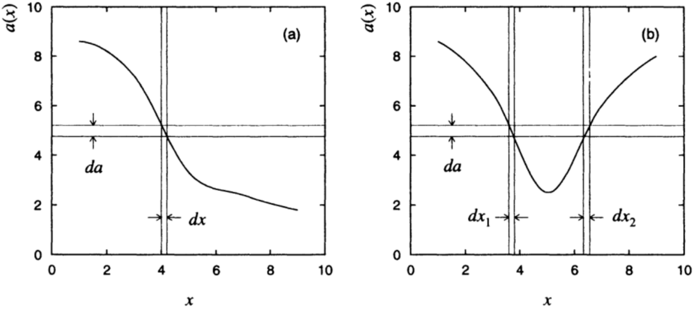
Therefore, the probability that she **does not** finish is, by the sum rule,

$$1 - P(n \leq 30|r,p) = 1 - \sum_{m=5}^{30} P(m|r,p) \approx 0.001$$

Table of Contents

Transformation of Variables

- It is often convenient to change variables when managing PDFs
- E.g., we have some $p(x|I)$ and we define $y = f(x)$, so we need to map $p(x|I)$ to $p(y|I)$



- Probability for x to occur between x and $x + dx$ must equal the probability for y to occur between y and $y + dy$

Transformation of Variables

- Consider a small interval δx around x' such that

$$p(x' - \frac{\delta x}{2} \leq x < x' + \frac{\delta x}{2} | I) \approx p(x = x' | I)\ \delta x$$

- $y = f(x)$ maps x' to $y' = f(x')$ and δx to δy. The range of y values in $y' \pm \delta y/2$ is equivalent to a variation in x between $x' \pm \delta x/2$, and so

$$p(x = x' | I)\ \delta x = p(y = y' | I)\ \delta y$$

  In the limit $\delta x \to 0$, this yields the PDF transformation rule

$$p(x|I) = p(y|I)\ \left|\frac{dy}{dx}\right|$$

Transformation of Variables

More than One Variable

- For more than one variable,

$$p(\{x_i\}|I)\ \delta x_1 \ldots \delta x_m = p(\{y_i\}|I)\ \delta^m \text{vol}(\{y_i\})$$

  where $\delta^m\text{vol}(\{y_i\})$ is an m-dimensional volume in y mapped out by the hypercube $\delta x_1 \ldots \delta x_m$

- The m-dimensional equivalent of the 1D transformation rule is

$$p(\{x_i\}|I) = p(\{y_i\}|I) \left|\frac{\partial(y_1, \ldots, y_m)}{\partial(x_1, \ldots, x_m)}\right|$$

  where the rightmost expression is the Jacobian matrix of partial derivatives dy_i/dx_j

Polar Coordinates

Example

For $x = R\cos\theta$ and $y = R\sin\theta$,

$$\left|\frac{\partial(x,y)}{\partial(R,\theta)}\right| = \begin{vmatrix}\cos\theta & -R\sin\theta \\ \sin\theta & R\cos\theta\end{vmatrix} = R(\cos^2\theta + \sin^2\theta) = R$$

Therefore, $p(R,\theta|I)$ is related to $p(x,y|I)$ by

$$p(R,\theta|I) = p(x,y|I)\cdot R$$

E.g., 2D Gaussian $\rightarrow$ Rayleigh distribution in R:

$$p(x,y|I) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} \implies p(R,\theta|I) = \frac{R}{2\pi\sigma^2}\exp\left\{-\frac{R^2}{2\sigma^2}\right\}$$

We have just equated the volume elements $dx\ dy = R\ dR\ d\theta$.

Applications

- We can imagine various situations in which these transformation rules are useful
- When measuring a quantity x, we can use the transformation to calculate the PDF of a derived quantity y (error propagation)
- In a problem of several variables $x, y, \ldots$, we might want to transform/rotate from coordinates with strong correlations to new variables $x', y', \ldots$ without correlations
- When sampling from a PDF it is quite convenient to transform from a PDF that is easy to generate to one that is more difficult
- We will discuss this in detail next class when we cover basic Monte Carlo techniques

Table of Contents

Generating Discrete Random Variables

- A probability generating function is a power series representation of the PMF of a discrete random variable
- These are not used in data analysis
- However, they are important in various branches of mathematics
- You may also see probability generating functions used in some calculations in statistical mechanics
- Note: the term is not universal, so your Stat Mech textbook may use such series but not refer to them as generating functions

Definition of a Generating Function

- Given a sequence of numbers $a_i : i = 0, 1, 2, \ldots$, the generating function of the sequence is defined as the power series

$$G(s) = \sum_{i=0}^{\infty} a_i s^i$$

for those values of s where the sum converges.
- For a given sequence, there exists a radius of convergence $R \geq 0$ s.t. the sum converges absolutely for $|s| < R$.
- $G(s)$ may be differential or integrated term by term any number of times when $|s| < R$.

Probability Generating Function

- Consider a count random variable $X \in \mathbb{N}$
- The probability that X is a given nonnegative integer k is

$$p_k = P(X = k), \quad k = 0, 1, 2, \ldots$$

- The probability generating function (PGF) of X is

$$G_X(s) = \sum_{k=0}^{\infty} p_k s^k = \mathrm{E}\left(s^X\right).$$

- Define: $G_X(0) = p_0$.
- Since $G_X(1) = 1$, the series converges absolutely for $|s| \leq 1$.

A Couple of Basic Properties

1. $G_X(0) = p_0 = P(X = 0)$.
2. $G_X(1) = P(X = 0) + P(X = 1) + P(X = 2) + \ldots = \sum_r P(X = r) = 1$.

Example

The generating function for a fair die is

$$G(1) = 0 + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$$

PGF of Constant Distribution

Example

Imagine if X is a constant or degenerate random variable – e.g., we roll a two-headed coin, or toss a die where all the faces are the same, so that

$$p_c = P(X = c) = 1,$$
$$p_k = 0 \text{ for } k \neq c.$$

In this case, the PGF of X is

$$G_X(s) = \mathrm{E}\left(s^X\right) = s^c.$$

PGF of Bernoulli and Binomial Trials

For a Bernoulli random variable which takes value 1 with probability p and value 0 with probability $q = 1 - p$,

$$\begin{aligned} p_0 &= 1 - p = q, \\ p_1 &= p, \\ p_k &= 0 \text{ if } k \neq 0 \text{ or } 1, \\ G_X(s) &= \mathrm{E}\left(s^X\right) = q + ps. \end{aligned}$$

For a binomial random variable X,

$$G_X(s) = (q + ps)^n.$$

For a Poisson random variable,

$$G_X(s) = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} s^k = e^{\lambda(s-1)}$$

Moments of the PGF

Given the PGF $G_X(s)$ we can obtain $p_k = P(X = k)$ in two ways:

1. Expand $G_X(s)$ in a power series and set $p_k =$ coefficient of s^k.
2. Differentiate $G_X(s)$ k times with respect to s and set $s = 0$.

The moments of a discrete random variable can be expressed in terms of the r^{th} derivative of $G_X(s)$ at $s = 1$. I.e.,

$$G_X^{(r)}(1) = \mathrm{E}\left[X(X-1)\ldots(X-r+1)\right]$$

Example: first two moments of X

$$\begin{aligned}
G_X^{(1)}(1) &= G_X'(1) = \mathrm{E}\left(X\right) \\
G_X^{(2)}(1) &= G_X''(1) = \mathrm{E}\left[X(X-1)\right] \\
&= \mathrm{E}\left(X^2\right) - \mathrm{E}\left(X\right) = \mathrm{var}\left(X\right) + \mathrm{E}\left(X\right)^2 - \mathrm{E}\left(X\right) \\
\therefore \mathrm{var}\left(X\right) &= G_X^{(2)}(1) - [G_X^{(1)}(1)]^2 + G_X^{(1)}(1).
\end{aligned}$$

Moments of the Poisson Distribution

Example

If X is a Poisson random variable, then

$$G_X(s) = e^{\lambda(s-1)}$$
$$G_X^{(1)}(s) = \lambda e^{\lambda(s-1)}$$
$$G_X^{(2)}(s) = \lambda^2 e^{\lambda(s-1)}$$

Therefore,

$$\mathrm{E}\left(X\right) = G_X^{(1)}(1) = \lambda e^0 = \lambda,$$
$$\mathrm{var}\left(X\right) = \lambda^2 - \lambda^2 + \lambda = \lambda.$$