Brief Treatise On Continuous Probability

John-Michael Laurel [1] `jm-laurel@berkeley.edu`

(1) Thinking Continuously

If we have a *continuous* random variable (r.v.) $\mathbf{X}$, then the range of $\mathbf{X}$ is uncountable. We now look for $\mathbf{X}$ falling into some interval $(x, x + dx)$, that is

$$\mathbb{P}(\mathbf{X} \in dx) = f_{\mathbf{X}}(x)dx$$

where $f_{\mathbf{X}}(x)$ is the *probability density function* (p.d.f.) of $\mathbf{X}$. The axioms of probability extend naturally. For non-negativity, $\mathbb{P}(\mathbf{X} \in dx) \geq 0$ implies

$$f_{\mathbf{X}}(x) \geq 0$$

Satisfying normalization, we integrate over the real number line

$$\int_{\mathbb{R}} f_{\mathbf{X}}(x)dx = 1$$

Lastly, for disjoint regions A and B

$$\int_{A \cup B} f_{\mathbf{X}}(x)dx = \int_A f_{\mathbf{X}}(x)dx + \int_B f_{\mathbf{X}}(x)dx$$

In the last axiom (additivity), integrating over A implies $\{x : x \in A\}$. Ditto for B. So (in most cases) you can think of A and B as non-overlapping segments on the real number line.

(2) Expectation and the Function Rule

Measures like *expectation* and *variance* extend as anticipated. The expectation of $g(\mathbf{X})$ is

$$\mathbb{E}\big[g(\mathbf{X})\big] = \int_{\mathbb{R}} g(x) f_{\mathbf{X}}(x)dx$$

and exists if and only if $-\infty < \mathbb{E}\big[g(\mathbf{X})\big] < \infty$.

[1] University of California, Berkeley (July 2019)
This document assumes understanding of basic probabilistic knowledge and discrete probability. See *Brief Treatise on Discrete Probability* (`Github Link`). Special thanks to Ella Hiesmayr for providing feedback on this document.

(3) Cumulative Distribution Function

The *Cumulative Distribution Function* (c.d.f.) completely characterizes a distribution and for say $\mathbf{X}$ is

$$\mathbb{P}(\mathbf{X} \leq x) = F_{\mathbf{X}}(x) = \int_{-\infty}^{x} f_{\mathbf{X}}(\tau)d\tau$$

We integrate with respect to a dummy variable to help distinguish from the upper bound of integration x. The c.d.f. is related to the p.d.f. by

$$\frac{d}{dx} F_{\mathbf{X}}(x) = f_{\mathbf{X}}(x)$$

In addition to being a strictly non-decreasing function, where $\lim_{x \to -\infty} F_{\mathbf{X}}(x) = 0$ and $\lim_{x \to \infty} F_{\mathbf{X}}(x) = 1$, suppose the inverse c.d.f. of $\mathbf{U} \sim \textbf{Uniform}(0, 1)$[2] denoted $F^{-1}(\mathbf{U})$, then $\mathbf{X} = F^{-1}(\mathbf{U})$ has distribution F.

(3.1) Minimums and Maximums

By employing the c.d.f., one can easily construct the distribution for the minimum and maximum of i.i.d. $\{\mathbf{X}_j\}_{j=1}^{n}$[3]. Let

$$\mathbf{X}_{\textbf{min}} = \textbf{min}\Big(\{\mathbf{X}_j\}_{j=1}^{n}\Big)$$

$$\mathbf{X}_{\textbf{max}} = \textbf{max}\Big(\{\mathbf{X}_j\}_{j=1}^{n}\Big)$$

, then

$$F_{\mathbf{X}_{\textbf{min}}}(x) = 1 - \prod_{j=1}^{n} \Big(1 - \mathbb{P}(\mathbf{X}_j \leq x)\Big)$$

$$F_{\mathbf{X}_{\textbf{max}}}(x) = \prod_{j=1}^{n} \mathbb{P}(\mathbf{X}_j \leq x)$$

The proof is rather straight forward and of course you could substitute $\mathbb{P}(\mathbf{X}_j \leq x)$ for $F_j(x)$ if you'd like.

(3.2) Expectation via CDF

You can compute $\mathbb{E}(\mathbf{X})$ for $\mathbf{X} \geq 0$ using its c.d.f.,

$$\mathbb{E}(\mathbf{X}) = \int_0^{\infty} \Big(1 - F(x)\Big)dx$$

$$= \int_0^{\infty} \mathbb{P}(\mathbf{X} > x)dx$$

Clearly this construction is the continuous extension of the tail-sum formula.

[2] refer to §(5.1) for the uniform distribution
[3] this notation is equivalent to $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}, \mathbf{X}_n\}$

(4) Change of Variable

Let $\mathbf{X}$ be a continuous r.v. with density $f_{\mathbf{X}}(x)$ and $\mathbf{Y} = g(\mathbf{X})$ have a derivative that is zero at only a finite number of points, then for $x = g^{-1}(y)$

$$f_{\mathbf{Y}}(y) = \sum_{\{x\,:\,g(x)=y\}} f_{\mathbf{X}}(x) \cdot \left|\frac{dy}{dx}\right|^{-1}$$

A fun example is finding the density of $\mathbf{Y} = \log(\mathbf{X})$, where $\mathbf{X} \sim$ **Uniform**$(0, 1)$.

(5) Named Distributions

This section highlights continuous distributions. For all values not defined on a p.d.f.'s domain, the p.d.f. assumes the value zero. More precisely, if $f_{\mathbf{X}}(x)$ is defined on some interval $\mathcal{I} \subseteq \mathbb{R}$, then $f_{\mathbf{X}}(x) = 0$ for all $x \notin \mathcal{I}$.

(5.1) Uniform Distribution

A uniform r.v. $\mathbf{X}$ on the interval (a, b), denoted $\mathbf{X} \sim$ **Uniform**(a, b) has density

$$f_{\mathbf{X}}(x) = \frac{1}{b-a}$$

$\mathbb{E}(\mathbf{X}) = \frac{a+b}{2}$, **var**$(\mathbf{X}) = \frac{(b-a)^2}{12}$

(5.2) Exponential Distribution

The continuous analog to $\mathbf{X} \sim$ **Geometric**(p). If $\mathbf{Y} \sim$ **Exponential**(λ), then

$$f_{\mathbf{Y}}(y) = \lambda e^{-\lambda y}$$

for $y \geq 0$ and $\lambda > 0$ a *rate*[4]. Observe that

$$p\mathbf{X} \to \textbf{Exponential}\,(1)$$

as $p \to 0$.[5] We interpret $\mathbf{Y}$ as the waiting time before some arrival. The *survival function*

$$\mathbb{P}(\mathbf{Y} > y) = e^{-\lambda y}$$

gives the probability that $\mathbf{Y}$ survives beyond y. In this context, $\mathbf{Y}$ is the waiting time until some end; usually death. $\mathbb{E}(\mathbf{Y}) = \frac{1}{\lambda}$, **var**$(\mathbf{Y}) = \frac{1}{\lambda^2}$

[4] synonyms: *hazard rate*, *failure rate*

[5] a scaled rendition of **Geometric**(p)

(5.2.1) Memoryless Property

Exponential(λ) is characterized by the *memoryless property*[6]. Framing $\mathbf{T} \sim$ **Exponential**(λ) as the lifetime of some entity, the distribution of the its remaining life after t has the same distribution of when time started. More formally,

$$\mathbb{P}(\mathbf{T} > t + s \,|\, \mathbf{T} > t) = \mathbb{P}(\mathbf{T} > s)$$

You can think of the remaining lifetime of a light-bulb as being exponentially distributed. It's chance of dying has the same distribution throughout its lifetime. In other words, it's utility is just as good as it was when it first turned on.

(5.2.2) Competing Exponentials

The continuous version of the Craps Principle. Let $\{\mathbf{X}_j\}_{j=1}^n \sim$ **Exponential**(λ_j) independent. Then

$$\mathbb{P}(\mathbf{X}_j < \mathbf{X}_1, \ldots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \ldots, \mathbf{X}_n) = \frac{\lambda_j}{\sum_{j=1}^n \lambda_j}$$

and when $\lambda_j = \lambda_i$ for all $1 \leq i, j \leq n$, then

$$\mathbb{P}(\mathbf{X}_j < \mathbf{X}_1, \ldots, \mathbf{X}_{j-1}, \mathbf{X}_{j+1}, \ldots, \mathbf{X}_n) = \frac{1}{n}$$

In the context of Craps, the $\mathbf{X}_j$'s are competing to see who arrives first.

(6) Gamma Distribution

A generalization[7] of **Exponential**(λ). Suppose we define $\{\mathbf{W}_i\}_{i=1}^r \sim$ **Exponential**(λ) independent and $\mathbf{T}_r = \sum_{i=1}^r \mathbf{W}_i$, then

$$f_{\mathbf{T}_r}(t) = \frac{\lambda^r}{(r-1)!}\, t^{r-1} e^{-\lambda t}$$

Showing the above requires a quick re-brief on **Poisson**(μ). Define $\mathbf{N}_t \sim$ **Poisson**(λt) as the number of arrivals in time t with rate $\lambda > 0$. It then follows,

$$\mathbb{P}(\mathbf{N}_t = k) = e^{-\lambda t}\, \frac{(\lambda t)^k}{k!}$$

We now have enough to derive $\mathbb{P}(\mathbf{T}_r \in dt)$.

[6] **Geometric**(p) also shares this property

[7] I think of it as daisy-chaining a bunch of exponentials

Proof. First observe $\mathbb{P}(\mathbf{T}_r \in dt)$ is equivalent to

$$\mathbb{P}(\mathbf{N}_t = r-1, \text{ arrival in } dt)$$

which for $\mathbf{W} \sim$ **Exponential** (λ) can be expressed as

$$\mathbb{P}(\mathbf{N}_t = r-1)\mathbb{P}(\mathbf{W} \in dt \,|\, \mathbf{N}_t = r-1)^8$$

The first factor is cake and the second is simply λdt. Stitching everything together,

$$\mathbb{P}(\mathbf{T}_r \in dt) = e^{-\lambda t}\,\frac{(\lambda t)^{r-1}}{(r-1)!} \times \lambda dt$$

∎

See **§(7)** for application. $\mathbb{E}(\mathbf{T}_r) = \frac{r}{\lambda}$, **var**$(\mathbf{T}_r) = \frac{r}{\lambda^2}$

(6.1) Gamma Function

The *Gamma function* makes an appearance in the density for $\mathbf{T}_r$ in **§(6)**. By definition of a density,

$$\int_0^\infty \frac{\lambda^r}{(r-1)!}\, t^{r-1}e^{-\lambda t}dt = 1$$

and for $\lambda = 1$,

$$(r-1)! = \int_0^\infty t^{r-1}e^{-t}dt$$

defines the Gamma function Γ. Notationally,

$$\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt$$

and for $r = n \in \mathbb{Z}^+$,

$$\Gamma(n) = (n-1)!$$

The two forms above highlight the fact that Gamma is not restricted to taking on positive integer values. e.g. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Defining gamma recursively,

$$\Gamma(n+1) = n\Gamma(n)$$

and is easily proved by induction.

(6.2) Beta Distribution and Order Statistics

We first introduce the idea of order statistics and proceed to a generalization of *uniform order statistics*, known as the *Beta Distribution*.

(6.2.1) Order Statistics

Let $\{\mathbf{X}_j\}_{j=1}^n$ be i.i.d. We then order the $\mathbf{X}_j$'s from smallest to largest

$$\mathbf{X}_{(1)} < \mathbf{X}_{(2)} < \cdots < \mathbf{X}_{(n-1)} < \mathbf{X}_{(n)}$$

and define $\mathbf{X}_{(k)}$ to be the k^{th} *largest order statistic*. It's prudent to note that the $\mathbf{X}_{(j)}$'s are *not* independent.[9] The k^{th} order statistic's density $f_{\mathbf{X}_{(k)}}(x)$ is defined for $x \in \mathbb{R}$ as

$$\binom{n}{k-1, 1, n-k} F^{k-1}(x)f(x)\Big[1 - F(x)\Big]^{n-k}$$

(6.2.2) Uniform Order Statistics

Suppose $\{\mathbf{X}_j\}_{j=1}^n \sim$ **Uniform**$(0,1)$ independent, then $f_{\mathbf{X}_{(k)}}(x)$ is clearly

$$f_{\mathbf{X}_{(k)}}(x) = \binom{n}{k-1, 1, n-k} x^{k-1}(1-x)^{n-k}$$

this is a nice segue into the Beta distribution.

(6.2.3) Beta Distribution

A generalization of independent uniform order statistics. We say $\mathbf{X} \sim$ **Beta**(α, β) if

$$f_{\mathbf{X}}(x) = \frac{1}{\mathbb{B}(\alpha,\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

, where $\mathbb{B}(\alpha,\beta) \in \mathbb{R}$ and is defined in terms of Γ

$$\mathbb{B}(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Observe that for $\alpha = k$ and $\beta = n-k+1$, we exactly have $f_{\mathbf{X}_{(k)}}(x)$, implying

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \binom{n}{k-1, 1, n-k}$$

When $\alpha, \beta \in \mathbb{Z}^+$, we can cast any Beta density as a joint distribution between independent uniform order statistics. $\mathbb{E}(\mathbf{X}) = \frac{\alpha}{\alpha+\beta}$, $\mathbb{E}\big[\mathbf{X}_{(k)}\big] = \frac{k}{n+1}$

[8] one could also write "$\mathbf{W} \in dt \,|\, W > t$" in the conditional

[9] e.g. If you know $\mathbf{X}_j = \mathbf{X}_{(1)}$, then the remaining $n-1$ random variables cannot possibly be the minimum of the original sequence of r.v.s.

(7) Poisson Arrival Process

We make use of notation in §**(6)**, contextualizing it to the *Poisson Arrival Process* which is characterized by 1. independent events, 2. constant average rate, and 3. no simultaneous hits. In the context of time and arrivals, we shrink each time interval such that each segment is a **Bernoulli**(p) trial.

(7.1) Arrival Epochs and Time Between Arrivals

Define a sequence of increasing r.v.s $\mathbf{T}_i$,

$$0 < \mathbf{T}_1 < \mathbf{T}_2 < \cdots < \mathbf{T}_{r-1} < \mathbf{T}_r$$

such that $\{\mathbf{T}_i\}_{i=1}^{r} \sim$ **Gamma**(i, λ) represents the time until the i^{th} arrival[10]. The *inter-arrival times* is a sequence of r.v. s

$$0 < \mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{r-2}, \mathbf{W}_{r-1}$$

where $\{\mathbf{W}_i\}_{i=1}^{r-1} \sim$ **Exponential** (λt) and describes the time between arrival epochs. Two equalities showcase how $\mathbf{T}_i$ and $\mathbf{W}_i$ are intertwined.

$$\mathbf{T}_r = \sum_{i=1}^{r} \mathbf{W}_i \quad \text{and} \quad \mathbf{W}_i = \mathbf{T}_i - \mathbf{T}_{i-1}$$

where in the second equality we assume $\mathbf{T}_0 = 0$.

(7.2) Counts of Arrivals

The number (or count) of arrivals up to time t, denoted $\mathbf{N}_t \sim$ **Poisson**(λt). For clarity, λ is a rate and $\mu = \lambda t$, is the *average expected arrivals* in a span of time t. One can recast λ into different quantities outside time, e.g. area, volume, etc.

(7.3) Logical Ramifications

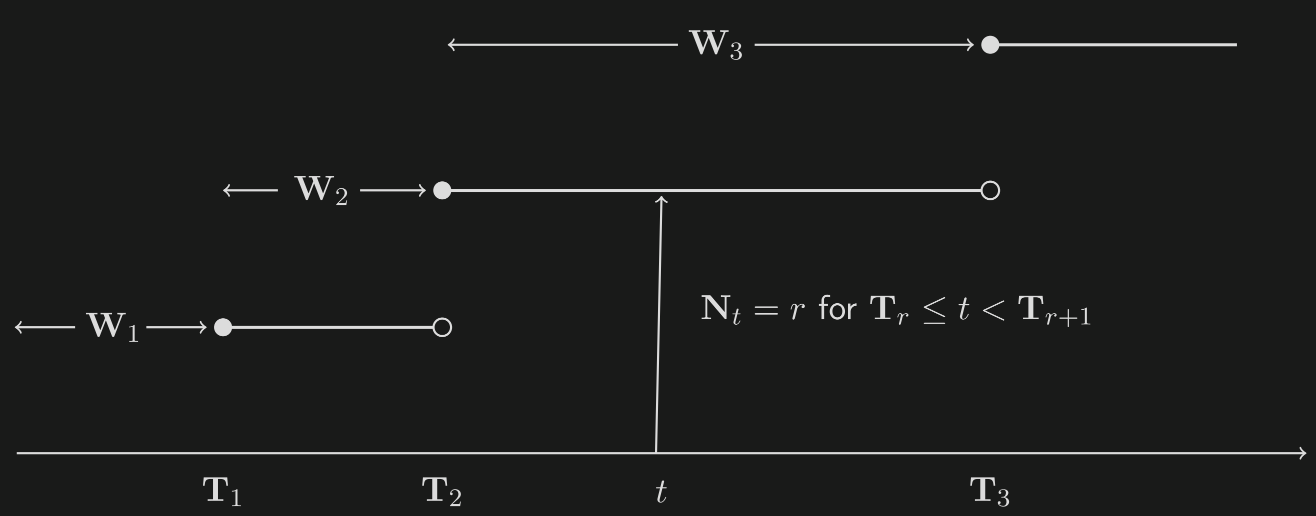
The following implications between r.v.s and events make sense.

$$\{\mathbf{T}_r > t\} = \{\mathbf{N}_t < r\} = \{\mathbf{N}_t \leq r - 1\}$$

In words, the r^{th} arrival arriving sometime after t is equivalent to having at most $r-1$ epochs arriving before time t.

(7.4) Diagram: Poisson Arrival Process

The digram below depicts a Poisson Arrival Process with arrivals $\{\mathbf{T}_i\}_{i=1}^{r}$, inter-arrival times $\{\mathbf{W}_i\}_{i=1}^{r-1}$, and counting process $\{\mathbf{N}_t \text{ for } t > 0\}$.



[10] I sometimes refer to $\mathbf{T}_i$ as the i^{th} *arrival epoch*.

(8) Normal Distribution

The continuous analog and approximation to **Binomial**(n, p) for n large and p not close to 0 or 1. If $\mathbf{Z} \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then

$$\phi_{\mathbf{Z}}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \textbf{exp}\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right]$$

The distribution's parameters μ and σ^2 are $\mathbb{E}(\mathbf{Z})$ and **var**$(\mathbf{Z})$ respectively.

(8.1) Standard Normal Distribution

When $\mu = 0$ and $\sigma^2 = 1$, we have the *standard normal distribution*. So if $\mathbf{X} \sim \mathcal{N}(0, 1)$, then

$$\phi_{\mathbf{X}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

We often transform a r.v. into a standard normal via a linear change of scale. So for $\mathbf{Y}$ not in standard units, we standardize $\mathbf{Y}$

$$\mathbf{Y}^* = \frac{1}{\sigma}\mathbf{Y} - \frac{\mu}{\sigma}$$

, hence $\mathbb{E}(\mathbf{Y}^*) = 0$ and **var**$(\mathbf{Y}^*) = 1$.

(8.2) Linear Combination of Normals

Let $\{\mathbf{X}_j\}_{j=1}^n \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right)$ independent and for $\alpha_j \in \mathbb{R}$, define $\mathbf{S}_n = \sum_{j=1}^n \alpha_j \mathbf{X}_j$, then

$$\mathbf{S}_n \sim \mathcal{N}\left(\sum_{j=1}^n \alpha_j \mu_j, \sum_{j=1}^n \alpha_j^2 \sigma_j^2\right)$$

That is the sum of normals is still normal. For the special case of $\{\mathbf{X}_j\}_{j=1}^n \sim \mathcal{N}(0, 1)$ independent and $\sum_{j=1}^n \alpha_j^2 = 1$, we know

$$\mathbf{S}_n \sim \mathcal{N}(0, 1)$$

by *rotational symmetry*. It can be easily verified that $\mathbb{E}(\mathbf{S}_n) = 0$ and **var**$(\mathbf{S}_n) = 1$. Parameterizing the constants in a linear combination of $\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(0, 1)$, namely $\mathbf{X}_\theta = \cos(\theta)\mathbf{X} + \sin(\theta)\mathbf{Y}$ showcases this idea. Now suppose $\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(0, 1)$ independent, then

$$\mathbf{X}^2 + \mathbf{Y}^2 \sim \textbf{Exponential}\left(\frac{1}{2}\right)$$

a miracle result.

(8.3) Rayleigh Distribution

A distribution with no parameters. Use when dealing with circles. Suppose $\mathbf{T} \sim$ **Exponential** $\left(\frac{1}{2}\right)$ and $\mathbf{R} = \sqrt{\mathbf{T}}$, then $\mathbf{R} \sim$ **Rayleigh** with density for $r \in \mathbb{R}^+$

$$f_R(r) = re^{-r^2}$$

and c.d.f.

$$F_R(r) = 1 - \frac{1}{2}e^{-r^2}$$

We also conclude for $\mathbf{X}$ and $\mathbf{Y}$ defined in **§(8.2)** that $\sqrt{\mathbf{X}^2 + \mathbf{Y}^2} \sim \mathcal{N}(0, 1)$.