CS70 In Simpler Terms - Note 13

written by Kevin Liu cs70.kevinmliu.com

August 2, 2017

1 Continuous Probability

Continuous probability is not that much different from discrete probability. Now, instead of dealing with the probability of distinct events, we are dealing with the probability of a range of events. This usually requires the use of integration.

There are 2 main functions that you need to be familiar with:

- Cumulative distribution function (CDF) - $F_x(x) = P(X \leqslant x)$
- Probability density function (PDF) - $f_x(x) = \frac{dF_x(x)}{dx}$
  - The PDF f must be nonnegative and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$
  - **Note:** Density $\neq$ probability. This implies that the density at any specific point can be greater than 1. So don't be surprised!

Recall that in discrete probability, the way to calculate the expectation of a r.v. is to take the summation of the probability of an event multiplied by the outcome of that event. Similarly in continuous probability, replace the summation with an integral:

- $E(x) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$
- $E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$

Tail Sum is an important trick for calculating the expectation of nonnegative r.v.:

- $E(x) = \int_0^{\infty} (1 - F_x(x))\,\mathrm{d}x = \int_0^{\infty} Pr(X > x)\,\mathrm{d}x$

2 Joint Distribution

- $Pr(a \leqslant X \leqslant b,\, c \leqslant Y \leqslant d) = \int_c^d \int_a^b f(x,y)\,\mathrm{d}x\,\mathrm{d}y$
  - $f(x,y) = f_x(x) f_y(y)$
  - f must be nonnegative, and $\int_{-\infty}^{\infty} \int_{\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$
- To obtain the marginal distribution of x:
  - $f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)\,\mathrm{d}y$ *(integrate over all values of y)*

- **Independent r.v.:**
  - $Pr(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) = Pr(a \leqslant X \leqslant b) * Pr(c \leqslant Y \leqslant d)$
  - Joint density of i.r.v. = product of marginal densities:
    $f(x, y) = f_x(x) * f_y(y)$

For the following distributions, I will not go through the derivations, but rather just list out the results, which you should be really familiar with.

3 Distributions and their Properties

- **Uniform Distribution**
  - $Unif[0, 1]$: $E(X) = \frac{1}{2}$, $Var(X) = \frac{1}{12}$
  - $Unif[a, b] : E(X) = \frac{a+b}{2}$, $Var(X) = \frac{1}{12}(b-a)^2$
- **Exponential Distribution**
  Analagous to the geometric distribution - *memoryless*
  The parameter $\lambda > 0$
  - $f(x) = \lambda e^{-\lambda x}$, $x \geqslant 0$
  - $F(x) = 1 - e^{-\lambda x} = Pr(X \leqslant x)$
  - $E(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$
  - **Important result:** $Pr(X > t) = \int_t^\infty \lambda e^{-\lambda x}\,\mathrm{d}x$

  Examples:

  1. Find the median of an exponential distribution:
     - $Pr(T \leqslant t) = \frac{1}{2} = Pr(T \geqslant t)$
     - $1 - e^{-\lambda t} = \frac{1}{2} \Rightarrow \frac{1}{2} = e^{-\lambda t} \Rightarrow t = \frac{\ln 2}{\lambda}$
  2. $Pr(min(X, Y) = X)$
     $= Pr(Y > X) = \int_0^\infty Pr(Y > y|X = x) f_x(x)\,\mathrm{d}x$
  3. Let $W = max\{X, Y\}$
     - $F_w(w) = Pr(W \leqslant w) = Pr(X \leqslant W) * Pr(Y \leqslant W) = F_x(w) * F_y(w)$
  4. Let $V = min\{X, Y\}$
     - $1 - F_v(v) = Pr(V > v) = Pr(X > v) * Pr(Y > v) = (1 - F_x(v)) * (1 - F_y(v))$
  5. Let $X_i \sim Exp(\lambda)$
     - $Z = min\{X_1, ...X_n\} \sim Exp(\lambda_1 + ...\lambda_n) \sim Exp(n\lambda)$
     - The sum is still exponential!
     - $E(Z) = \frac{1}{n\lambda}$
- **Normal/Gaussian Distribution**
  - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}$

- $\mathbb{P}[X \leqslant a] = \mathbb{P}\big[Y \leqslant \frac{a-\mu}{\sigma}\big]$, where Y is standard normal a.k.a. $Y = \frac{X-\mu}{\sigma}$
  * From the above, we get $X = \sigma Y + \mu$
  * $Pr(a \leqslant Y \leqslant b) = Pr(\sigma a + \mu \leqslant X \leqslant \sigma b + \mu)$
- Standard normal $= \mu = 0$, $\sigma = 1$
- $68 - 95 - 99.7$ Rule: 68% of events in a normal distribution are within 1 s.d. (± 1), 95% of events are within 2 s.d. (± 2), and 99.7% are within 3 s.d..
- Let $Z = aX + bY$, where $X \sim (\mu_x, \sigma_x^2)$, $Y \sim (\mu_y, \sigma_y^2)$ and X, Y are independent
  * $\mu_z = a\mu_x + b\mu_y$
  * $\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2$