More Probability and Monte Carlo Integration

yesterday: def of probability consepts, events,
prob, random variables, inverse cdf method

Gallery of probability distributions

1) $X \sim \text{Bernoulli}(p)$; $P(X=1) = p$; $P(X=0) = 1-p$
$\text{Bin}(1,p)$

$$E(X) = 1 \cdot p + 0(1-p) = p$$
$$Var(X) = p(1-p)$$

2) $X \sim \text{Bin}(n,p)$; $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

$$E(X) = \underbrace{p + \dots + p}_{n} = np$$
$$Var(X) = np(1-p)$$

3) $X \sim \text{Geometric}_1(p)$ - # of Bernoulli trials to get one success

$P(X=k) = (1-p)^{k-1} p$, $k = 1, 2, 3, \dots$ $\quad \underbrace{00001}_{5}$

$X \sim \text{Geometric}_2(p)$ - # of Bernoulli failers to get one success

$P(X=k) = (1-p)^k p$ $\quad \underbrace{0000}_{4}1$

4) $X \sim \text{Poisson}(\lambda)$

$$P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots$$

$$\boxed{E(X) = Var(X) = \lambda}$$ - unrealistic assumption in many applications

$\uparrow$ very strong mean-variance relationship

5) $X \sim \text{NegBin}(r,p)$ - negative binomial

# of failures in independent Bernoulli trials before before a pre-defined number of successes (r)

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

$$E(X) = \frac{r(1-p)}{p} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

6) $X \sim N(\mu, \sigma^2)$ $\qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$E(X) = \mu$; $Var(X) = \sigma^2$

7) $X \sim Exp(\lambda)$ $\qquad \boxed{f(x) = \lambda e^{-\lambda x} \cdot 1\{x \geq 0\}}$

$E(X) = \frac{1}{\lambda}$; $Var(X) = \frac{1}{\lambda^2}$ $\qquad \lambda > 0$

8) $X \sim Gamma(\alpha, \beta)$ $\qquad \boxed{f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \cdot 1\{x \geq 0\}}$

$E(X) = \frac{\alpha}{\beta}$; $Var(X) = \frac{\alpha}{\beta^2}$

If $\alpha = 1 \Rightarrow X \sim Gamma(1, \beta) = Exp(\beta)$

9) $X \sim Beta(\alpha, \beta)$ $\qquad f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$

$x \in (0, 1)$

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

10) $p \sim Beta(\alpha, \beta) \Rightarrow X | p \sim Bin(n, p)$ – this results in a Beta-Binomial distribution

The above process is called compounding.

Moreover, Neg Bin can be constructed as:

$\lambda \sim \text{Gamma}(\alpha, \beta) \Rightarrow X|\lambda \sim \text{Poisson}(\lambda)$

Strong Law of Large Numbers (SLLN)

Let $X_1, X_2, \dots$ be independent and identically distributed (iid) random variables with

$$\mu = E(X_1) < \infty. \text{ Then}$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$

SLLN says that empirical averages of iid random variables converge to the theoretical average/expectation

Central Limit Theorem (CLT)

Let $X_1, X_2, \dots$ be independent and identically distributed (iid) random variables with

$\mu = E(X_1) < \infty$ and $0 < \sigma^2 = Var(X_1) < \infty$

and let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \sim N(0,1) \text{ for large } n \text{ approximately}$$

Informally, CLT say that for large n, the empirical average $\bar{X}_n$ behaves as $N(\mu, \sigma^2/n)$

Scaling of the variance by $1/n$ implies that averaging reduces variability, which is intuitive.

Monte Carlo Integration

Objective: $E(h(X)) = \int h(x) f(x)\,dx$, where X is a random variable with probability density function $f(x)$

or

$E(h(X)) = \sum_{k=1}^{\infty} h(x_k)\, p_k$, where X is a discrete random variable with prob mass function $p_1, p_2, \ldots$

If $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ and $E(h(X_1)) < \infty$, then SLLN $\Rightarrow$ $\frac{1}{n} \sum_{i=1}^{n} h(X_i) \to E[h(X_1)]$

$$\bar{h}_n = \frac{1}{n} \sum_{i=1}^{n} h(X_i) \approx E[h(X_1)]$$

$$Var(\bar{h}_n) = Var\left(\frac{1}{n} \sum_{i=1}^{n} h(X_i)\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} h(X_i)\right) =$$

$$= [\text{independence}] = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{Var(h(X_i))}_{Var(h(X_1))} =$$

$$= \frac{1}{n^2} \times n \times Var[h(X_1)] =$$

$$= \frac{1}{n} Var[h(X_1)] \approx \frac{1}{n} \times \underbrace{\frac{1}{n-1} \sum_{i=1}^{n} [h(X_i) - \bar{h}_n]^2}_{\text{sample variance } v_n}$$

where v_n is the sample variance of $h(X_1), \ldots, h(X_n)$

Moreover, CLT: $\bar{h}_n \overset{\text{approximately}}{\sim} N\left(\mu, \frac{\sigma_n}{n}\right)$

We can form 95% confidence interval:

$\bar{h}_n \pm 1.96 \sqrt{\frac{\sigma_n}{n}}$ – Monte Carlo error

Importance Sampling

Objective: $E_f[h(X)] = \int h(x) f(x) dx$

We cannot or don't want to sample from the target density $f(x)$. Instead, we want to sample from some other, perhaps simpler, distribution with density $g(x)$.

$$E_f[h(x)] = \int h(x) f(x) dx = \int \underbrace{\boxed{h(x) \frac{f(x)}{g(x)}}}_{z(x)} g(x) dx =$$

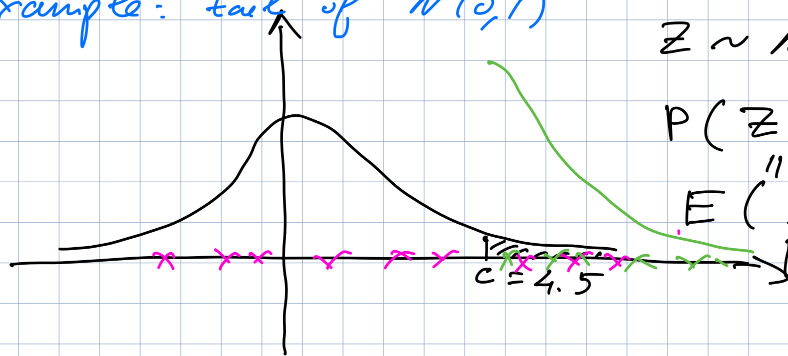
$$= E_g\left[h(y) \frac{f(y)}{g(y)}\right], \text{ where } y \sim g(y)$$

This derivation suggests that we can generate iid $y_1, \dots, y_n \overset{iid}{\sim} g(x)$ and use SLLN to arrive at the approximation

$$E_f[h(x)] \approx \frac{1}{n} \sum_{i=1}^{n} \underbrace{\boxed{\frac{f(y_i)}{g(y_i)}}}_{w_i} h(y_i)$$

w_i – importance sampling weights

Example: tail of $N(0,1)$



$Z \sim N(0,1)$

$P(Z>c)$ - objective

$= E(\underbrace{1_{\{Z>c\}}}_{h(Z)})$

Naive Monte Carlo: $Z_1, \dots, Z_n \overset{iid}{\sim} N(0,1)$.

Then
$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Z_i > c\}} \approx E(1_{\{Z>c\}}) = P(Z>c)$$

Importance sampling: Simulate

$$y_1, \dots, y_n \overset{iid}{\sim} \text{Shifted-Exp}(c, 1)$$

Shifted exponential density is

$$g(y) = e^{-(y-c)} 1_{\{y>c\}}$$

$N(0,1)$ density

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{\varphi(y_i)}{g(y_i)} \cancel{1_{\{y_i > c\}}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\varphi(y_i)}{g(y_i)}$$