

MSA101/MVE187 2017 Lecture 4

Petter Mostad

Chalmers University

September 7, 2017

Monte Carlo Integration

- ▶ We want to estimate (compute) an integral (given a r.v. X)

$$I = \Pr(f(X) \leq \alpha) = \int I(f(x) \leq \alpha)\pi(x) dx = \int g(x)\pi(x) dx = E(g(X))$$

- ▶ We want to do it computing an average:
 - ▶ Simulate x_1, \dots, x_m from $\pi(x)$.
 - ▶ Compute

$$\hat{I}_m = \frac{1}{m} (g(x_1) + \dots + g(x_m))$$

- ▶ We can often easily generate lots of data, i.e., m is very large.
- ▶ We use the Central Limit Theorem, to approximate that, as $m \rightarrow \infty$,

$$\hat{I}_m \sim \text{Normal}(I, \text{Var}(g(X))/m)$$

as long as the first two moments of $g(X)$ exists.

Monte Carlo Integration, cont.

- ▶ We can estimate $\text{Var}(g(X))$ with

$$\text{Var}(g(X)) \approx s^2 = \frac{1}{m-1} \sum_{i=1}^m \left(g(x_i) - \hat{l}_m \right)^2$$

- ▶ With this, we can estimate a 95% confidence interval for l with the sample variance

$$\hat{l}_m \pm 1.96s/\sqrt{m}$$

with a similar interpretation as usual.

- ▶ A possibility is to compute and plot the estimate and the confidence interval as a function of m : See Example 3.3 in Robert.

Example: Estimating a proportion

In our main example above, we have $g(X) = I(f(X) \leq \alpha)$, and we want to estimate $p = E(I(f(X) \leq \alpha))$.

- ▶ Then

$$\text{Var}(I(f(X) \leq \alpha)) = E(I(f(X) \leq \alpha)) - E(I(f(X) \leq \alpha))^2 = p - p^2.$$

- ▶ Thus the accuracy of estimates is proportional to $s = \sqrt{p(1-p)}$.
- ▶ The accuracy seems to improve when $p \rightarrow 0$, but what matters is the *relative* accuracy,

$$\sqrt{p(1-p)}/p = \sqrt{1/p - 1}$$

which is bad when $p \rightarrow 0$.

- ▶ In other words: Estimating a tail quantile from a probability distribution by counting the number of times sampled values are in the tail is not very efficient.

Approximating quantiles by simulation

To compute an approximate interval containing, e.g., 90% of the probability for a random variable X :

- ▶ Simulate x_1, \dots, x_n from X .
- ▶ Order them by size and find the 5'th and 95'th empirical percentile.
- ▶ In R, use, e.g., `quantile(..)`.

Simulation from a uniform distribution

- ▶ Simulation from $\text{Uniform}[0, 1]$ is the basis of all computer based simulation.
- ▶ What does it mean that $x_1, \dots, x_n \sim \text{Uniform}[0, 1]$ is "random"? A possible interpretation: We have no way to predict the coming numbers; the best guess for their distribution is $\text{Uniform}[0, 1]$.
- ▶ The computer uses a deterministic function applied to a seed ("pseudo-random"). The seed can be set (in R with `set.seed(...)`) or is taken from the computer clock.
- ▶ It should be in practice impossible to apply any kind of visualiation or compute any kind of statistic which has properties other than those predicted when the sequence x_1, \dots, x_n is *iid* $\text{Uniform}[0, 1]$.

Simulating from discrete distributions

- ▶ If X is a random variable on a finite set of real numbers, the cumulative distribution can be computed in a vector. X can be simulated by comparing a uniform random variable U to the numbers in this vector. Example: Binomial distribution.
- ▶ If X is a random variable on a countable set of real numbers, one can describe a finite set that contains approximately all the probability and do as above. Example: The Poisson distribution.

The inverse transform

- ▶ Let X be a random variable with invertible cumulative distribution function $F(x)$. If $U \sim \text{Uniform}[0, 1]$, then $F^{-1}(U)$ is a random sample from X .
- ▶ Note:

$$P(F^{-1}(u) \leq \alpha) = Pr(F(F^{-1}(u)) \leq F(\alpha)) = Pr(u \leq F(\alpha)) = F(\alpha)$$

- ▶ Example: The exponential distribution $\text{Exp}(\lambda)$ has density $\pi(X) = \lambda \exp(-x\lambda)$ and cumulative distribution

$$F(x) = 1 - \exp(-\lambda x)$$

$F(x) = u$ gives $F^{-1}(u) = -1/\lambda \log(1 - u)$. As $1 - u$ is also uniform, we can simulate with

$$-1/\lambda \log(u)$$

The inverse transform, cont.

- ▶ Example: Logistic distribution. Best defined by defining its cumulative distribution (for standard logistic distribution):

$$F(x) = 1/(1 + \exp(-x))$$

Easy to invert. The distribution can be adjusted with changing the mean and the scale, in a standard way.

- ▶ Example: Cauchy distribution. Density:

$$\pi(x) = 1/(\pi(1 + x^2)).$$

The cumulative distribution is

$$F(x) = 1/2 + 1/\pi \arctan(x)$$

Easy to invert.

Transforming samples

- ▶ Example: One can prove that, if X_1, \dots, X_n is a random sample from $\text{Exp}(1)$ then

$$\beta \sum_{i=1}^n X_i \sim \text{Gamma}(n, \beta)$$

- ▶ Example: One can prove that, if X_1, \dots, X_n is a random sample from $\text{Exp}(1)$ then

$$\frac{\sum_{i=1}^a X_i}{\sum_{i=1}^{a+b} X_i} \sim \text{Beta}(a, b).$$

- ▶ Example: One can prove that, if U_1, U_2 is a random sample from $\text{Uniform}[0, 1]$, then

$$\left(\sqrt{-2 \log(U_1)} \cos(2\pi U_2), \sqrt{-2 \log(U_1)} \sin(2\pi U_2) \right)$$

is a random sample from $\text{Normal}(0, 1)$.

Transformation of random variables

- ▶ Recall from basic probability theory: If $f(x)$ is a density function, and $x = h(y)$ is a monotone transformation, then the density function for y is

$$f(h(y))|h'(y)|$$

- ▶ If we apply the INVERSE of h on an variable with known density, we get the density of the resulting variable using the formula above.
- ▶ Example application: The non-informative prior for the precision τ of a Normal distribution is the *improper* distribution with "density" $\pi(\tau) \propto 1/\tau$. We have that $\tau = h(\sigma^2) = 1/\sigma^2$. We have that, when $h(x) = 1/x$, $h'(x) = -1/x^2$. Thus the non-informative prior for the variance σ^2 of a normal distribution is given as

$$\pi(\sigma^2) \propto \frac{1}{1/\sigma^2} \left| -\frac{1}{(\sigma^2)^2} \right| = \frac{1}{\sigma^2}$$

Transformation of multivariate random variables

- ▶ If x is a vector, if $f(x)$ is a multivariate density function, and if $x = h(y)$ is a bijective differentiable transformation, then the multivariate density function for y is

$$f(h(y))|J(y)|$$

where $|J(y)|$ is the determinant of the Jacobian matrix for the vector function $h(y)$.

- ▶ One application of this is to prove the identity used above to simulate from the normal distribution.

The multivariate normal

- ▶ $x \sim \text{Normal}_k(\mu, \Sigma)$ if

$$\pi(x) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^t \Sigma^{-1}(x - \mu)\right)$$

- ▶ NOTE: If x_1, \dots, x_k are i.i.d $\text{Normal}(0, 1)$ then $x = (x_1, \dots, x_n)^t \sim \text{Normal}_k(0, I)$.
- ▶ If $x \sim \text{Normal}_k(0, I)$ then $Ax \sim \text{Normal}(0, AA^t)$.
- ▶ THUS: To simulate from $\text{Normal}(\mu, \Sigma)$:
 - ▶ Simulate k independent standard normal random variables into a vector x .
 - ▶ Compute the (lower triangular) Choleski decomposition S of Σ : We then have that $\Sigma = SS^t$.
 - ▶ Compute $Sx + \mu$: It is multivariate normal, and has the right expectation and variance matrix.

Simulating the marginal distribution

- ▶ Generally: If you have a sample $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ from a joint distribution of X and Y , then x_1, x_2, \dots, x_n is a sample from the marginal distribution of X .
- ▶ Simple application: If $\tau \sim \text{Gamma}(k/2, 1/2)$ and $x | \tau \sim \text{Normal}(0, 1/\tau)$, then the marginal distribution of x is a Student t-distribution with k degrees of freedom. To simulate:
 - ▶ Draw τ from $\text{Gamma}(k/2, 1/2)$.
 - ▶ Then draw x from $\text{Normal}(0, 1/\tau)$.
- ▶ Much more generally: To simulate for example from the predictive distribution for x_{NEW} in a Bayesian model, simulate from the joint distribution with density $\pi(x_{NEW}, \theta | x)$, where x is the data and θ is the parameters. Then take the coordinates of the sample pertaining to x_{NEW} .

Rejection sampling

- ▶ Sometimes we cannot easily simulate from a density $f(x)$, (the "target density") but we *can* simulate from an "instrumental" density $g(x)$ that approximates $f(x)$.
- ▶ If we can find a constant M such that $f(x)/g(x) \leq M$ for all x (and if f and g have the same support), we can use *rejection sampling* to sample from f :
 - ▶ Sample X using $g(x)$.
 - ▶ Draw u uniformly on $[0, 1]$.
 - ▶ If $u \cdot M \leq f(x)/g(x)$ accept x as a sample, otherwise reject x and start again.

Rejection sampling, cont.

- ▶ NOTE: Applicable in any dimension.
- ▶ The acceptance rate is $1/M$, so we want to use a small M .
- ▶ NOTE: We may in fact do this with $f(x)$ and $g(x)$ equal to the densities up to a constant, still a valid method!
- ▶ NOTE: When $g(x)$ integrates to 1, the integral of $f(x)$ can be approximated as the acceptance rate multiplied by M .
- ▶ Example: Random variables with log-concave densities can be simulated with this method.