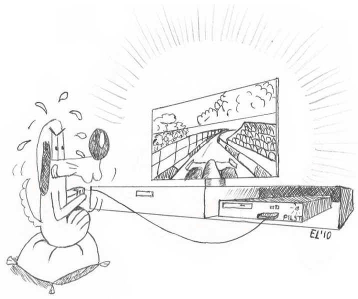
PERFORMANCE EVALUATION EXERCISES

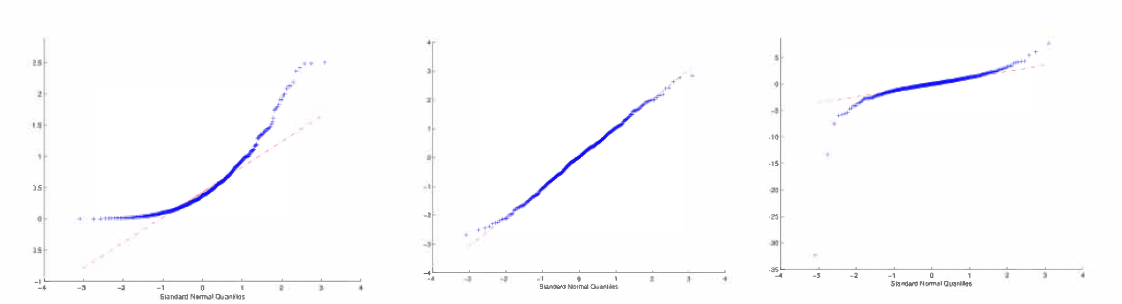
SIMULATION

With Solutions Jean-Yves Le Boudec, August 23, 2022



In the text below, rand$(1, n)$ produces a sequence of n independent samples of the uniform distribution between 0 and 1 and randn$(1, n)$ produces n independent samples of the standard normal distribution.

1. Which of the following three normal qq-plots is for an exponential distribution ?



(a) ☒ The first

(b) ☐ The second

(c) ☐ The third

(d) ☐ The first and second

(e) ☐ The first and third

(f) ☐ The second and third

(g) ☐ All

(h) ☐ None

Solution. The positive part of the tail of the exponential distribution is heavier than for the normal distribution because the pdf decays in e^{-x} which is much less fast than $e^{-x^2/2}$. The negative positive part of the tail of the exponential distribution is lighter than for the normal distribution because the exponential never has negative values. Only the first plot has these features.

2. Independent simulation outputs are obtained by...

A executing the runs on parallel threads using the same seed.

B executing the runs on parallel threads using truly random seeds.

C using the last RNG state of one run as seed to the next run.

(a) ☐ A

(b) ☐ B

(c) ☐ C

(d) ☐ A or B

(e) ☐ A or C

(f) ☒ B or C

(g) ☐ All

(h) ☐ None

3. The random variable X is integer and its distribution is given by:

$$\begin{aligned}\mathbb{P}(X=-1) &= 0.1 \\ \mathbb{P}(X=0) &= 0.8 \\ \mathbb{P}(X=1) &= 0.1\end{aligned}$$

Write the pseudo-code of a simulation program to generate a sample of X.

Solution.

```
U ← rand(1,1)
if U ≤ 0.1
      return (−1)
else if U ≤ 0.9
      return (0)
else
      return (1)
```

4. The standard Pareto distribution with index 1 has PDF $f_X(x) = \frac{1}{x^2}\mathbf{1}_{\{x>1\}}$. Write the pseudo-code of a simulation program to generate a sample of X.

Solution. The CDF can be computed in closed form. For $c > 1$:

$$\begin{aligned}&\text{for } c>1: \quad && F(c) = \mathbb{P}(X \le c) = \int_1^c \frac{1}{x^2}dx = \left[-\frac{1}{x}\right]_1^c = 1 - \frac{1}{c} \\ &\text{for } c \le 1: \quad && F(c) = 0\end{aligned}$$

It can easily be inverted. For $0 < p < 1$, the equation $F(x) = p$ can be solved in closed form:

$$1 - \frac{1}{x} = p \Leftrightarrow x = \frac{1}{1-p}$$

i.e. $F^{-1}(p) = \frac{1}{1-p}$. Thus, if U is a uniform random variable between 0 and 1, $\frac{1}{1-U}$ has the required distribution. Note that U and $1-U$ have the same distribution, thus $\frac{1}{U}$ also has the required distribution. This gives the following pseudo-code:

```
return (1/rand(1,1))
```

Note that here we are assuming that the random number generator $\text{rand}(1,1)$ never returns 0, which is true for matlab. If it were not true, you would to catch the case where it returns 0.

5. We draw a (X, Y) point uniformly at random in the unit disk. We observe its radius R and its angle Θ.

(a) Compute the PDFs $f_{R,\Theta}$, f_F and f_Θ.

(b) Are R and Θ uniform ? Independent ?

Solution.

(a) We can use the textbook formula that says that if the random vector $\vec{X}$ is mapped to a new random vector $\vec{Y} = h(\vec{X})$ then the PDF of the old random vector $\vec{X}$ is equal to the PDF of the new random vector $\vec{Y}$ multiplied by the absolute value of the Jacobian of the transformation $J_h(x) = \det(\nabla h_{|x})$:

$$f_{\vec{X}}(x) = f_{\vec{Y}}(h(x)) \times |J_h(x)|$$

Here we take as mapping h

$$(x,y) \mapsto (r,\theta) \text{ with } \begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \end{cases}$$

with $\theta \in [0, 2\pi)$ and $0 \le r \le 1$. We can take R, Θ as old vector and X, Y as new vector. The jacobian of the transformation is

$$J(r,\theta) = \det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} = r\left(\cos^2(\theta) + r\sin^2(\theta)\right) = r$$

thus

$$f_{R,\Theta}(r,\theta) = r f_{X,Y}(x,y) \text{ with } \begin{cases} x = r\cos(\theta) \\ y = r\sin(\theta) \end{cases}$$

Now X, Y is uniform in the unit disk, therefore its PDF is

$$f_{X,Y}(x,y) = \frac{1}{\pi}\mathbf{1}_{\{x^2+y^2\le 1\}}$$

because π is the area of the unit-disk. Furthermore the condition $x^2 + y^2 \le 1$ is equivalent to $\rho \le 1$. Thus:

$$f_{R,\Theta}(r,\theta) = \frac{r}{\pi}\mathbf{1}_{\{0\le r\le 1\}}\mathbf{1}_{\{0\le\theta<2\pi\}}$$

The PDF of R is obtained by integrating over θ, we obtain

$$f_R(r) = 2r\mathbf{1}_{\{0\le r\le 1\}}$$

and similarly

$$f_\Theta(\theta) = \frac{1}{2\pi}\mathbf{1}_{\{0\le\theta<2\pi\}}$$

(b) The PDF of the angle Θ is constant therefore it is uniform. In contrast, the PDF of the radius is not constant (but is proportional to r) therefore the radius is not uniformly distributed: a larger radius is more likely.

The joint PDF has product form, therefore the radius and angle are independent. Knowing the radius of a point in the unit disk gives no information about its angle (and vice versa).

6. The function $H()$ returns a positive random number with PDF $f()$ and CDF $F()$; c is a positive constant such that $F(c) < 1$. The following program computes a random variable Y.

```
do forever
    X ← H()
    if X > c
        Y ← X
        return (Y)
```

Does the program always terminate ? Under these conditions, what is the distribution of Y? What are its PDF and CDF ? How many iterations does the program take in average?

Solution. The program implements rejection sampling by first sampling X from $F()$ then testing whether $X > c$. The probability that one iteration succeeds is $p = \mathbb{P}(X > c) = 1 - F(c) > 0$. The program terminates because $p > 0$; the number of iterations minus one follows a geometric distribution with parameter p. The mean number of iterations is thus $1/p$.

The distribution of Y is the conditional distribution of X given that $X > c$. Note that obviously $f_Y(x) = 0$ if $x \le c$ since, by construction, Y is always larger than c. Now let $x > c$; we have

$$\mathbb{P}(Y \in [x, x+dx]) = f_Y(x)dx + o(dx) = \frac{\mathbb{P}(X \in [x,x+dx] \text{ and } X > c)}{\mathbb{P}(X > c)}$$

Since we assume $x > c$ the event ($X \in [x, x+dx]$ and $X > c$) is equal to ($X \in [x, x+dx]$) thus

$$\mathbb{P}(Y \in [x, x+dx]) = \frac{\mathbb{P}(X \in [x,x+dx])}{\mathbb{P}(X > c)} = \frac{f(x)dx + o(dx)}{1 - F(c)}$$

and finally

$$\begin{array}{lll} f_Y(x) & = \frac{f(x)}{1-F(c)} & \text{if } x > c \\ f_Y(x) & = 0 & \text{if } x \le c \end{array}$$

The CDF follows by integration: for $x \le c$, $F_Y(x) = 0$ and for $x > c$

$$F_Y(x) = \int_c^x f_Y(x)dx = \frac{F(x) - F(c)}{1 - F(c)}$$

7. The time T between successive sending of messages by a sensor is random but always larger than some constant τ. Its average is $\bar{t}$ (with $\bar{t} > \tau$).

   (a) Bart models T as follows:
   $$T = \tau + X, \quad X \sim \text{Exp}(\lambda_1)^1$$
   What are the PDF and CDF of T ? What value should λ_1 have ? Write the pseudo-code of a simulation program to generate a sample of T according to Bart. How many calls to the random number generator does this program use ? Run your program to generate a histogram of 1000 samples; which shape do you expect ?

   (b) Lisa models T as a sample of the conditional distribution of Y given $Y > \tau$ with $Y \sim \text{Exp}(\lambda_2)$. What are the PDF and CDF of T ? What value should λ_2 have ? Write the pseudo-code of a simulation program to generate a sample of T using rejection sampling according to Lisa's model. How many calls to the random number generator does this program use ?

   (c) Compare these modelling and simulation methods.

Solution.

(a) The PDF of T is given by the rule of change of variable: if $Y = h(X)$ then the PDF of X is equal to $f_X(x) = f_Y(h(x)) |h'(x)|$. Here:

$$\begin{aligned} f_T(t) &= f_X(t-\tau) = \lambda_1 e^{-\lambda_1(t-\tau)} \mathbf{1}_{\{t \ge \tau\}} \\ &= \lambda_1 e^{\lambda_1 \tau} e^{-\lambda_1 t} \mathbf{1}_{\{t \ge \tau\}} \end{aligned}$$

[1] $\text{Exp}(\lambda)$ stands for the exponential distribution with rate λ.

The PDF is a shifted exponential. The CDF is also derived: if if $Y = h(X)$ and h is increasing then $F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(h(X) \le (h(x)) = \mathbb{P}(Y \le h(x)) = F_Y(h(x))$. Here:

$$\begin{aligned} F_T(t) &= F_X(t-\tau) = \left(1 - e^{-\lambda_1(t-\tau)}\right) \mathbf{1}_{\{t\ge\tau\}} \\ &= \left(1 - e^{\lambda_1\tau} e^{-\lambda_1 t}\right) \mathbf{1}_{\{t\ge\tau\}} \end{aligned}$$

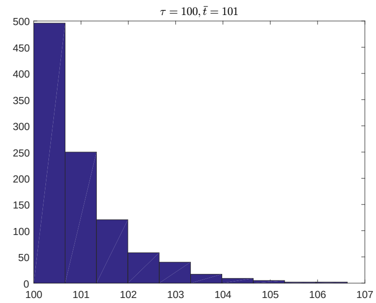
The expectation of T should be $\bar{t}$; the mean of $\text{Exp}(\lambda)$ is $\frac{1}{\lambda}$. Thus we must have $\frac{1}{\lambda_1} + \tau = \bar{t}$, i.e.

$$\lambda_1 = (\bar{t} - \tau)^{-1}$$

A simulation program for T is derived from the method for generating a sample of the uniform distribution, and is

$$\text{return}(\tau - (\bar{t} - \tau) \log \text{rand}(1,1))$$

It uses one call per sample.
The histogram is an approximation of the PDF, so it should be similar to an exponential decreasing curve, starting at τ.



1000 samples out of Bart's simulation, $\tau = 100$, $\bar{t} = 101$.

(b) The conditional distribution (see other exercise) has PDF

$$f_T(t) = \frac{f_Y(t)\mathbf{1}_{\{t>\tau\}}}{1 - F_Y(\tau)} = \frac{\lambda_2 e^{-\lambda_2 t}\mathbf{1}_{\{t>\tau\}}}{e^{-\lambda_2\tau}} = \lambda_2 e^{\lambda_2\tau} e^{-\lambda_2 t}\mathbf{1}_{\{t>\tau\}}$$

It has the same form as Bart's, so the distributions are identical. The condition is therefore $\lambda_2 = \lambda_1 = (\bar{t} - \tau)^{-1}$.
The rejection sampling method according to Lisa would be:

do forever
  $Y \leftarrow -(\bar{t} - \tau) \log \text{rand}(1,1)$
  if $Y > \tau$
  return(Y)

Bart's method is more efficient as it calls the RNG only once per output. In contrast, Lisa's method uses in average $\frac{1}{1-F_Y(\tau)} = e^{\frac{\tau}{\bar{t}-\tau}}$ calls to the RNG. This is always more than 1 and much more if $\bar{t} - \tau$ is small compared to τ.

8. The following simulation program computes a random vector (X_1, X_2):

$X_1 \leftarrow \text{randn}(1,1)$
$B \leftarrow \text{rand}(1,1)$
  if $B > 0.5$
    $X_2 \leftarrow X_1$
  else
    $X_2 \leftarrow -X_1$
return(X_1, X_2)

(a) What is the distribution of X_1 ? of X_2 ?

(b) What is the distribution of $X_1 + X_2$? What are its PDF and CDF ?

(c) Plot 1000 samples of (X_1, X_2), what do you see ? Compare to a plot of 1000 samples of (Y_1, Y_2) where Y_1 and Y_2 are independent and standard normal. Is (X_1, X_2) a gaussian random vector ?

Solution.

(a) X_1 is standard gaussian by construction.

Solution 1 The conditional distribution of X_2 given $B > 0.5$ is standard gaussian. The conditional distribution of X_2 given $B \leq 0.5$ is the opposite of a standard gaussian. Because the standard gaussian distribution is symmetric around 0, the conditional distribution of X_2 given $B \leq 0.5$ is also standard gaussian. The conditional distribution of X_2 given B is always the same, independent of the value taken by B, and is standard gaussian; therefore the distribution of X_2 is standard gaussian (furthermore, X_2 and B are independent !)

Solution 2 To compute the distribution of X_2 let us write $\mathbb{E}(\varphi(X_2))$ for any (bounded) test function φ. We have

$$\mathbb{E}\left(\varphi(X_2)|B > 0.5\right) = \mathbb{E}\left(\varphi(X_1)\right)$$

Also:

$$\mathbb{E}\left(\varphi(X_2)|B \leq 0.5\right) = \mathbb{E}\left(\varphi(-X_1)\right) = \mathbb{E}\left(\varphi(X_1)\right)$$

because the standard gaussian distribution is symmetric around 0, i.e. X_1 and $-X_1$ have the same distribution. By the law of total probabilities

$$\begin{aligned}
\mathbb{E}\left(\varphi(X_2)\right) &= \mathbb{E}\left(\varphi(X_2)|B < 0.5\right)\mathbb{P}(B < 0.5) + \mathbb{E}\left(\varphi(X_2)|B \leq 0.5\right)\mathbb{P}(B \leq 0.5) \\
&= \mathbb{E}\left(\varphi(X_1)\right)\left(\mathbb{P}(B < 0.5) + \mathbb{P}(B \leq 0.5)\right) \\
&= \mathbb{E}\left(\varphi(X_1)\right)
\end{aligned}$$

This is true for any test function φ thus X_2 is a standard gaussian random variable.

(b) Given that $B > 0.5$, we have $X_1 + X_2 = 2X_1$, so the conditional distribution of $X_1 + X_2$ given that $B > 0.5$ is gaussian with 0 mean and variance equal to 4. Given that $B \leq 0.5$, we have $X_1 + X_2 = 0$, so the conditional distribution of $X_1 + X_2$ given that $B \leq 0.5$ is a constant equal to 0 (also called "Dirac"). By the law of total probabilities, we have, for any (bounded) test function φ:

$$\mathbb{E}\left(\varphi(X_1 + X_2)\right) = 0.5\int_{-\infty}^{+\infty}\varphi(x)f_{N_{0,4}}(x)dx + 0.5\varphi(0) \quad (1)$$

where $f_{N_{0,4}}()$ is the PDF of the normal distribution with mean 0 and variance 4.

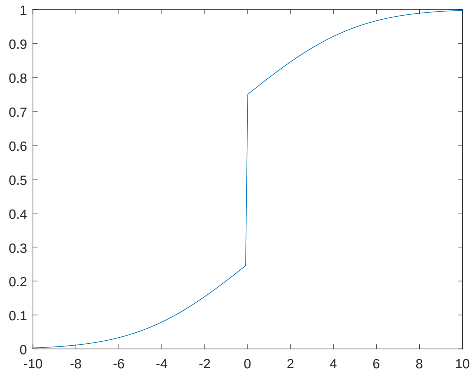
The Dirac distribution does not have a PDF *stricto sensu* and the same holds for $X_1 + X_2$; but we often pretend it has, using the dirac impulse function δ, which has the (magical) property that $\int_{-\infty}^{+\infty}\varphi(x)\delta(x)dx = \varphi(0)$ for any test function φ. With this notation, the PDF of $X_1 + X_2$ is

$$f_{X_1+X_2}(x) = 0.5f_{N_{0,4}}(x) + 0.5\delta(x)$$

The CDF of $X_1 + X_2$ is well defined even stricto sensu; we obtain it by letting $\varphi(x) = \mathbf{1}_{\{x \leq c\}}$ in (1):

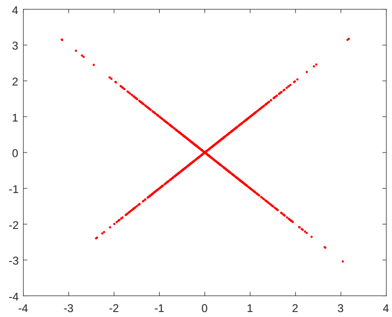
$$\begin{aligned}
F_{X_1+X_2}(c) &= \int_{-\infty}^{+\infty}\mathbf{1}_{\{x \leq c\}}f_{N_{0,4}}(x)dx + 0.5\cdot\mathbf{1}_{\{0 \leq c\}} \\
&= 0.5F_{N_{0,4}}(c) + 0.5\cdot\mathbf{1}_{\{0 \leq c\}}
\end{aligned}$$

where $F_{N_{0,4}}()$ is the PDF of the normal distribution with mean 0 and variance 4.
Clearly, $X_1 + X_2$ is not gaussian (for example because the CDF is discontinous at 0).

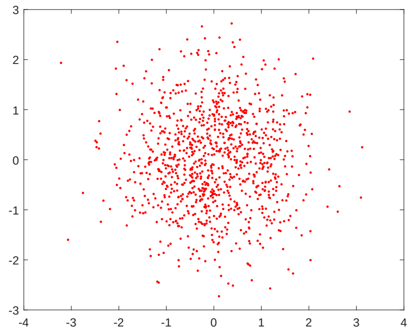


CDF of $X_1 + X_2$.

(c) For (X_1, X_2) we see 1000 points distributed on the two diagonals. Clearly, (X_1, X_2) is not a gaussian vector (even though each coordinate is gaussian) because if it were, $X_1 + X_2$ would be gaussian. In contrast, (Y_1, Y_2) is (by definition) a gaussian random vector.



1000 samples of (X_1, X_2).



1000 samples of (Y_1, Y_2).

9. The random variable X is nonnegative and has complementary CDF $F^c(x) = \mathbb{P}(X > c) = (1 + x)e^{-x}1_{x>0}$. What is the PDF of X ? Write the pseudo-code of a simulation program to generate a sample of X.

Solution. The PDF is the derivative of the CDF thus

$$f_X(x) = -\frac{d}{dx}F_X^c(x) = xe^{-x}$$

CDF inversion is not easy because F is hard to invert. We can use rejection sampling. We need to find a distribution with PDF f_Y such that it is easy to sample, Y is defined on $\mathbb{R}^+$ and f_X/f_Y is bounded. There are many such choices, for example an exponential distribution with rate $\lambda = 0.5$

$$f_Y(x) = 0.5e^{-0.5x}$$

We have

$$\frac{f_X(x)}{f_Y(x)} = 2xe^{-0.5x}$$

By studying the derivative of this expression we find that it is maximum when $x = 2$ hence

$$\frac{f_X(x)}{f_Y(x)} \leq 4e^{-1}$$

A rejection sampling method for X is thus as follows.

```
c = 4e^{-1}
do forever
        (U, V) ← rand(1, 2)
        V ← c · V
        Y ← −2 log(U)                                   // Sample Y ∼ Exp(0.5)
        if V ≤ 2Y e^{−0.5Y}
                return(Y)
```

An equivalent, simpler form is

```
do forever
        (U, V) ← rand(1, 2)
        if V ≤ −e log(U)U
                return(−2 log(U))
```

10. We want to simulate a random point (X, Y) in the unit disk, whose PDF is

$$f_{X,Y}(x, y) = \eta\sqrt{x^4 + y^4}\mathbf{1}_{\{x^2+y^2\leq 1\}}$$

In the formula,η is a normalizing constant. Give the pseudo-code of a program that produces a sample of (X, Y). Try it and sample 10'000 points. How does this visually compare to the uniform distribution on the unit the disk ?

Solution. We can use rejection sampling by comparing the pdf to that of a uniform point in the square $[-1; +1] \times [-1; +1]$, whose pdf is $f_S(x, y) = 1/4$ for $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$. We need to find a bound c on the ratio of densities:

$$\frac{f^n_{X,Y}(x, y)}{f_S(x, y)} = \frac{\sqrt{x^4 + y^4}\mathbf{1}_{\{x^2+y^2\leq 1\}}}{1/4} \leq 4\sqrt{2} = c$$

With rejection sampling, we draw $U \sim \text{Unif}(0, c)$ and accept the sample (X, Y) if $U \leq \frac{f^n_{X,Y}(x,y)}{f_S(x,y)}$, which is equivalent to

$$\left(X^2 + Y^2 \leq 1\right) \text{ and } \left(\frac{U}{4} \leq \sqrt{X^4 + Y^4}\right)$$

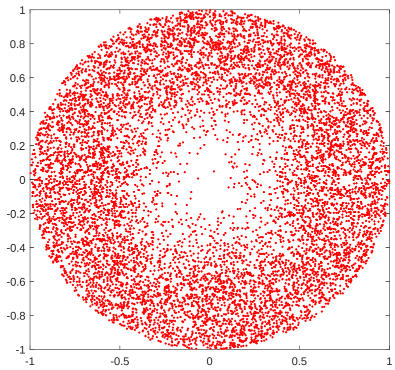
This gives the program:

```
c = √2
do forever
      (X, Y, U) ← rand(1, 3)
      X ← 2X − 1; Y ← 2Y − 1
      U ← c · U
      if X^2 + Y^2 ≤ 1
            if U ≤ 4√(X^4 + Y^4)
                  return (X, Y)
```

which can be optimized a bit in order to avoid unnecessary calls to the RNG:

$c = \sqrt{2}$
do forever
    $(X, Y) \leftarrow \text{rand}(1, 2)$
    $X \leftarrow 2X - 1; Y \leftarrow 2Y - 1$
    if $X^2 + Y^2 \leq 1$
        $U \leftarrow c \cdot \text{rand}(1, 1)$
        if $U \leq 4\sqrt{X^4 + Y^4}$
            return (X, Y)

Visually, the distribution gives fewer points towards the center than the uniform distribution, which is expected because when X [resp. Y] is small, X^4 [resp. Y^4] is very small.



10'000 samples of (X, Y).

11. What does this program compute ? (A is a subset of $[0; 1]^n$).

$N \leftarrow 0$
do $r = 1 : R$
    if $\text{rand}(n, 1) \in A$
        $N \leftarrow N + 1$
return (N/R)

Solution. A Monte-Carlo estimate of the volume of A.

It computes a Monte-Carlo estimate of

$$\mathbb{E}\left(\mathbf{1}_{\{X \in A\}}\right) = \mathbb{P}(X \in A)$$

where X is uniformly distributed in $[0; 1]^n$. Now the PDF of X is 1 over $[0; 1]^n$ and 0 outside. Hence

$$\mathbb{P}(X \in A) = \int_{[0;1]^n} \mathbf{1}_{\{X \in A\}} dx = \int_A dx = \text{volume}(A)$$