Problem 1. Sample from Cauchy [Bi] Ex.11.3 Given a random variable z uniformly distributed over $(0,1)$, find a transformation $y = f(\mathrm{z})$ such that y has Cauchy distribution

$$p_{\mathrm{y}}(y) = \frac{1}{\pi}\frac{1}{1+y^2}.$$

Solution. Note that

$$\int \frac{1}{a^2+u^2}\,\mathrm{d}u = \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + C$$

We need

$$z = h(y) = \int_{-\infty}^{y} p_{\mathrm{y}}(y)\,\mathrm{d}y = \frac{1}{\pi}\tan^{-1}(y) + \frac{1}{2}$$

Therefore

$$y = h^{-1}(z) = \tan\left(\pi\left(z - \frac{1}{2}\right)\right)$$

■

Problem 2. Box-Muller [Bi] Ex.11.4 Suppose z_1 and z_2 are uniformly distributed over the unit circle (disk). Show that

$$y_1 = z_1\left(\frac{-2\ln r^2}{r^2}\right)^{1/2}, \quad y_2 = z_2\left(\frac{-2\ln r^2}{r^2}\right)^{1/2}$$

where $r = z_1^2 + z_2^2$, has the joint density

$$p_{(y_1,y_2)}(y_1, y_2) = \left[\frac{1}{\sqrt{2\pi}}\exp\left(-y_1^2/2\right)\right]\left[\frac{1}{\sqrt{2\pi}}\exp\left(-y_2^2/2\right)\right]$$

Solution. We know that

$$p(y_1, y_2) = p(z_1, z_2)\left|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}\right|$$

To find the Jacobian, we use the polar coordinate as intermediate and apply chain role. We define polar coordinate as

$$\begin{aligned} \theta &= \tan^{-1}\frac{z_2}{z_1} \\ r^2 &= z_1^2 + z_2^2 \end{aligned} \quad \text{and,} \quad \begin{aligned} z_1 &= r\cos\theta \\ z_2 &= r\sin\theta \end{aligned}$$

Using the polar coordinate, we have

$$\frac{\partial(z_1, z_2)}{\partial(r,\theta)} = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} \quad \left|\frac{\partial(z_1, z_2)}{\partial(r,\theta)}\right| = r\left(\cos^2\theta + \sin^2\theta\right) = r.$$

We can represent y as

$$y_1 = z_1\left(\frac{-2\ln r^2}{r^2}\right)^{1/2} = \left(-2\ln r^2\right)^{1/2}\cos\theta \tag{1}$$

$$y_2 = z_2\left(\frac{-2\ln r^2}{r^2}\right)^{1/2} = \left(-2\ln r^2\right)^{1/2}\sin\theta \tag{2}$$

and thus

$$\frac{\partial(y_1, y_2)}{\partial(r,\theta)} = \begin{pmatrix} -2\cos\theta\left(-2\ln r^2\right)^{-1/2} r^{-1} & -2\sin\theta\left(-2\ln r^2\right)^{-1/2} r^{-1} \\ -\sin\theta\left(-2\ln r^2\right)^{1/2} & \cos\theta\left(-2\ln r^2\right)^{1/2} \end{pmatrix}$$

$$\left|\frac{\partial(r,\theta)}{\partial(y_1, y_2)}\right| = \left|\frac{\partial(y_1, y_2)}{\partial(r,\theta)}\right|^{-1} = \left(-2r^{-1}\left(\cos^2\theta + \sin^2\theta\right)\right)^{-1} = -\frac{r}{2}.$$

Applying the chain role, we have

$$\left|\frac{\partial\left(z_1, z_2\right)}{\partial\left(y_1, y_2\right)}\right|=\left|\frac{\partial\left(z_1, z_2\right)}{\partial(r, \theta)} \frac{\partial(r, \theta)}{\partial\left(y_1, y_2\right)}\right|=\left|\frac{\partial\left(z_1, z_2\right)}{\partial(r, \theta)}\right|\left|\frac{\partial(r, \theta)}{\partial\left(y_1, y_2\right)}\right|=-\frac{r^2}{2}$$

We will only use the absolute value of this.

By squaring both side of (1) and (2) and adding them together, we have

$$y_1^2+y_2^2=-2 \ln r^2 \quad \Rightarrow \quad r^2=\exp \left(-\frac{y_1^2+y_2^2}{2}\right)$$

Since $\left(z_1, z_2\right)$ is uniform, we have $p\left(z_1, z_2\right)=\frac{1}{\pi}$. Finally,

$$p(y_1, y_2)=p(z_1, z_2)\left|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}\right|=\frac{1}{\pi} \frac{r^2}{2}=\frac{1}{2 \pi} \exp \left(-\frac{y_1^2+y_2^2}{2}\right)$$

■

Problem 3. Gibbs sampling Consider the Gibbs sampler for a vector of parameters $\boldsymbol{x}=\left(x_1, \cdots, x_M\right)^{\top}$. Suppose at the s-th step $\boldsymbol{x}^{(s)}$ is sampled from the target distribution $p(\boldsymbol{x})$ and then $\boldsymbol{x}^{(s+1)}$ is generated using the Gibbs sampler. Show that the marginal probability $P\left(\boldsymbol{x}^{(s+1)} \in \mathbb{A}\right)$ equals the target distribution $\int_{\mathbb{A}} p(\boldsymbol{x}) d \boldsymbol{x}$.

Solution. We can write $\boldsymbol{x}^{(s+1)}$ as,

$$p(\boldsymbol{x}^{(s+1)})=p(x_i^{(s+1)} \mid \boldsymbol{x}_{-i}^{(s+1)}) p(\boldsymbol{x}_{-i}^{(s+1)})=p(x_i^{(s+1)} \mid \boldsymbol{x}_{-i}^{(s)}) p(\boldsymbol{x}_{-i}^{(s)})=\frac{p(x_i^{(s+1)} \mid \boldsymbol{x}_{-i}^{(s)})}{p(x_i^{(s)} \mid \boldsymbol{x}_{-i}^{(s)})} p(\boldsymbol{x}^{(s)})$$

Therefore, the marginal probability converges to the target distribution. ■

Problem 4. Entropy Recall that the entropy of a discrete random variable X is defined to be

$$H(X)=-\sum_{x \in \mathbb{X}} p(x) \log_2 p(x)$$

where $\mathbb{X}$ is the set of all possible values of X.

1. A fair coin is flipped until the first head occurs. Let X denote the number of flips required. Find the entropy $H(X)$ in bits.

   Solution. Since $X=n$ means that first $n-1$ flips are tail and last flip is head. Suppose the probability of a head is p. Then, we have

$$P(X=n)=(1-p)^{n-1}(p)^{-1}$$

   Thus, the entropy is

$$\begin{aligned}
H(X) & =-\sum_{n=1}^{\infty}(1-p)^{n-1} p \log \left((1-p)^{n-1} p\right) \\
& =-\left[\sum_{n=1}^{\infty}(1-p)^{n-1} p \log p+\sum_{n=1}^{\infty}(1-p)^{n-1} p \log (1-p)^{n-1}\right] \\
& =-\left[\sum_{m=0}^{\infty}(1-p)^m p \log p+\sum_{m=0}^{\infty} m(1-p)^m p \log (1-p)\right] \\
& =\frac{-p \log 0}{1-(1-p)}-\frac{p(1-p) \log (1-p)}{p^2} \\
& =\frac{-p \log p-(1-p) \log (1-p)}{p}
\end{aligned}$$

Here we have a fair coin, $p = 1/2$, and the entropy is $H(X) = 2$ bits. ■

2. What is the relationship of $H(X)$ and $H(Y)$ if $Y = 2^X$?

Solution. Suppose $y = f(x)$, then

$$p(y) = \sum_{x:y=f(x)} p(x)$$
$$p(x) \leq \sum_{x:y=f(x)} p(x) = p(y)$$

Thus,

$$\sum_{x:y=f(x)} p(x)\log p(x) \leq p(y)\log p(y)$$

Then, we have

$$\begin{aligned} H(X) &= -\sum_{x\in X} p(x)\log p(x) \\ &= -\sum_{y\in Y}\sum_{x:y=f(x)} p(x)\log p(x) \\ &\geq -\sum_{y\in Y} p(y)\log p(y) = H(Y) \end{aligned}$$

It is equal if and only if $f(x)$ is one to one.

Since $Y = 2^X$ is one to one, $H(Y) = H(X)$ ■

Problem 5. Differential entropy Calculate the (differential) entropy of the following.

1. The exponential density $p(x) = \lambda e^{-\lambda x}, x \geq 0$.

Solution.

$$\begin{aligned} h(x) &= -\lambda\int_0^\infty e^{-\lambda x}\ln\left(\lambda e^{-\lambda x}\right)\mathrm{d}x \\ &= \lambda\int_0^\infty e^{-\lambda x}\ln\left(\frac{1}{\lambda}e^{\lambda x}\right)\mathrm{d}x \\ &= -\ln(\lambda)\lambda\int_0^\infty e^{-\lambda x}\mathrm{d}x + \lambda^2\int_0^\infty xe^{-\lambda x}\mathrm{d}x \\ &= -\ln(\lambda)\lambda\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty + \lambda^2\left(\left[-\frac{1}{\lambda}xe^{-\lambda x}\right]_0^\infty + \frac{1}{\lambda}\int_0^\infty e^{-\lambda x}\mathrm{d}x\right) \\ &= -\ln(\lambda) + \lambda\left[-\frac{1}{\lambda}e^{-\lambda x}\right]_0^\infty \\ &= 1 - \ln(\lambda) \end{aligned}$$

■

2. The sum of x_1 and x_2 where x_1 is independent from x_2 and $p_{x_i}(x) = \mathcal{N}\left(x \mid \mu_i, \sigma_i^2\right)$ for $i = 1, 2$.

Solution. We know that the sum of two independent Gaussian distributions is still a Gaussian distribution (wiki). Let $y = x_1 + x_2$, we have

$$p_{x_1}(x) = \mathcal{N}\left(x \mid \mu_1, \sigma_1^2\right), \quad p_{x_2}(x) = \mathcal{N}\left(x \mid \mu_2, \sigma_2^2\right), \quad p_y(y) = \mathcal{N}\left(y \mid \mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

Then, for the differential entropy of Gaussian distribution, we have

$$
\begin{aligned}
h(X) &= -\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right)\mathrm{d}x \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\ln\left(\sigma\sqrt{2\pi}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)\right)\mathrm{d}x \\
&= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-t^2\right)\ln\left(\sigma\sqrt{2\pi}\exp\left(t^2\right)\right)\mathrm{d}t \quad \left(\text{substituting } t=\frac{x-\mu}{\sqrt{2}\sigma}\right) \\
&= \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left(\ln(\sigma\sqrt{2\pi})+\ln\left(\exp\left(t^2\right)\right)\right)\exp\left(-t^2\right)\mathrm{d}t \\
&= \frac{\ln(\sigma\sqrt{2\pi})}{\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left(-t^2\right)\mathrm{d}t+\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}t^2\exp\left(-t^2\right)\mathrm{d}t \\
&= \frac{\sqrt{\pi}\ln(\sigma\sqrt{2\pi})}{\sqrt{\pi}}+\frac{1}{\sqrt{\pi}}\left(\left[-\frac{t}{2}\exp\left(-t^2\right)\right]_{-\infty}^{\infty}+\frac{1}{2}\int_{-\infty}^{\infty}\exp\left(-t^2\right)\mathrm{d}t\right) \\
&= \ln(\sigma\sqrt{2\pi})+\frac{1}{2\sqrt{\pi}}\int_{-\infty}^{\infty}\exp\left(-t^2\right)\mathrm{d}t \\
&= \ln(\sigma\sqrt{2\pi})+\frac{\sqrt{\pi}}{2\sqrt{\pi}} \\
&= \ln(\sigma\sqrt{2\pi})+\frac{1}{2}
\end{aligned}
$$

Therefore, $h(Y)=\ln\left(\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}\right)+\frac{1}{2}$ ■

Problem 6. Change of variable

Recall that $H(\boldsymbol{x})=-\int p_{\boldsymbol{x}}(\boldsymbol{x})\ln p_{\boldsymbol{x}}(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x}$. Prove:

$$H(\boldsymbol{A}\boldsymbol{x})=\ln|\det(\boldsymbol{A})|+H(\boldsymbol{x}).$$

Solution. Recall that when we make a change of variables, the probability density is transformed by the Jacobian of the change of variables.

$$p(\boldsymbol{x})=p(\boldsymbol{y})\left|\frac{\partial y_i}{\partial x_j}\right|=p(\boldsymbol{y})\det\boldsymbol{A}$$

Then the entropy of $\boldsymbol{y}$ is

$$H(\boldsymbol{y})=-\int p(\boldsymbol{y})\ln p(\boldsymbol{y})\ \mathrm{d}\boldsymbol{y}=-\int p(\boldsymbol{x})\ln\left(p(\boldsymbol{x})\det(\boldsymbol{A})^{-1}\right)\mathrm{d}\boldsymbol{x}=H(\boldsymbol{x})+\ln|\det(\boldsymbol{A})|$$

■