

CT5202: Photorealistic Rendering

Density Estimation for Rendering Techniques

Lecturer: Bochang Moon

Density Estimation

- $P(a < X < b) = \int_a^b f(x)dx$ for all $a < b$
- Problem:
 - Input: a set of observed data points generated from an unknown probability density function (pdf)
 - Output: an estimate of the pdf
- A parametric approach:
 - We assume the pdf is a known parametric family of distributions
 - e.g., if we assume the samples are drawn from a normal distribution, we can estimate the mean μ and variance σ^2 with a sample mean and sample variance

Density Estimation

- In practice,
 - It is hard to employ the parametric approach as our target function (e.g., radiance) may be very complex
- Nonparametric approaches:
 - Histograms
 - Kernel estimator
 - Nearest neighbor method

Histograms

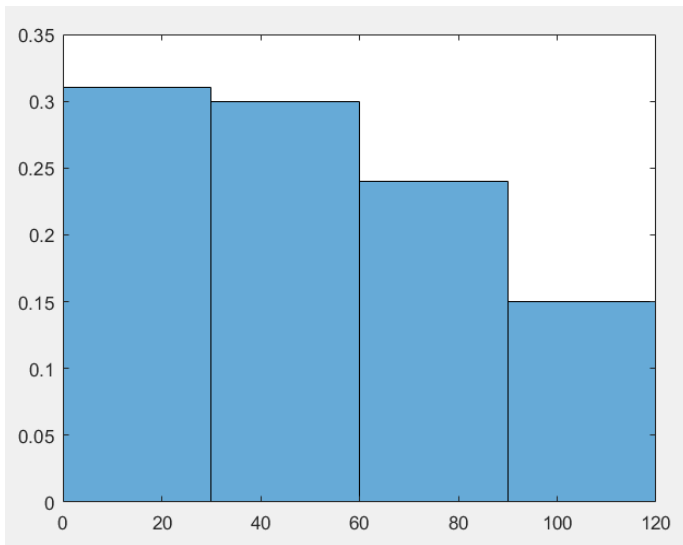
- Input: n real data X_1, \dots, X_n
- Given an origin x_0 and a bin width h , the bins of histograms are defined by the intervals $[x_0 + mh, x_0 + (m + 1)h)$, where m is an integer
- $$\hat{f}(x) = \frac{\text{\# of } X_i \text{ in same bin as } x}{nh}$$

Data

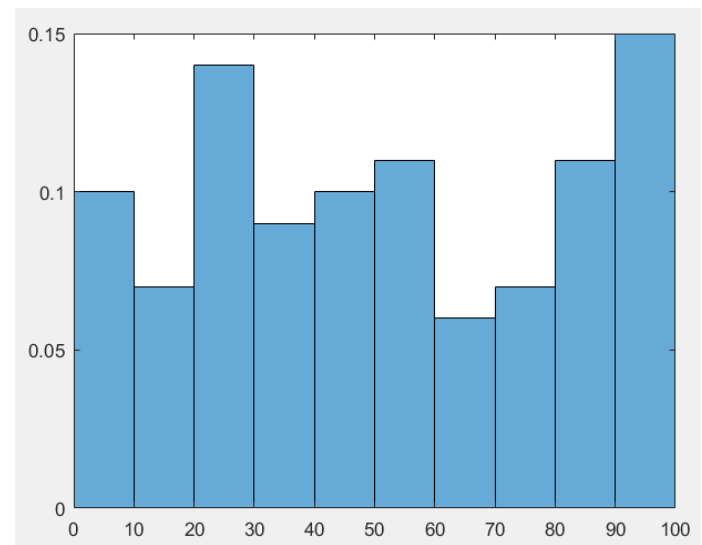
- Example: 100 random samples from a unknown pdf

75	95	32	96	97	25	5	20	47	95
34	10	99	39	39	21	20	49	26	58
48	93	15	55	97	80	14	26	70	85
86	84	68	8	78	17	86	85	55	74
82	52	88	90	4	46	46	46	11	92
56	42	27	70	31	29	26	50	25	7
86	83	39	21	30	61	76	87	90	50
67	17	1	23	50	93	38	42	8	32
9	42	23	8	67	62	53	94	15	90
4	55	4	49	99	99	55	68	25	73

Histograms



$h=30$



$h=10$

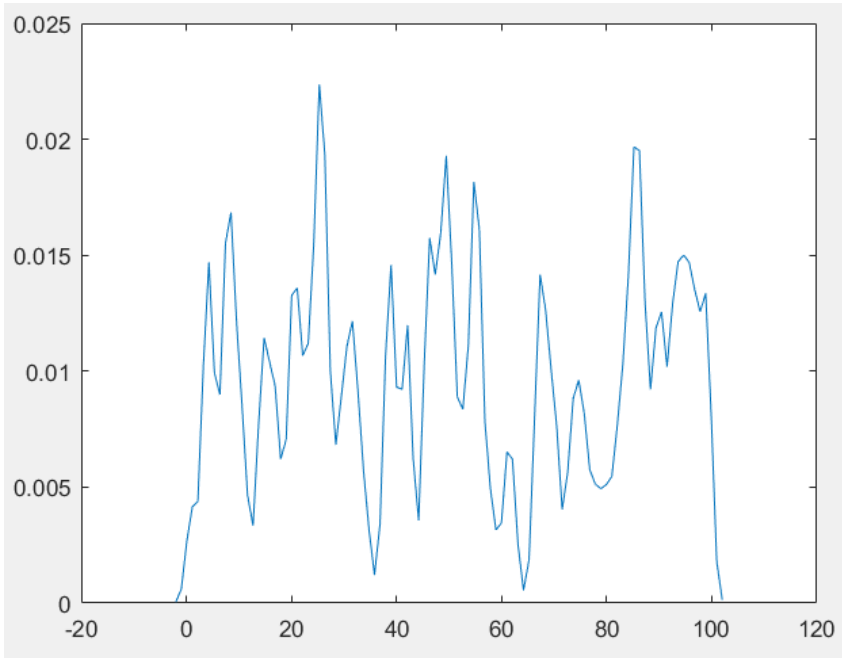
Histograms

- It discretizes the density function
- The parameter h controls the amount of smoothing
- The number of bins grows exponentially as the dimensionality of the data increases
- The density function has discontinuities at the bin boundaries

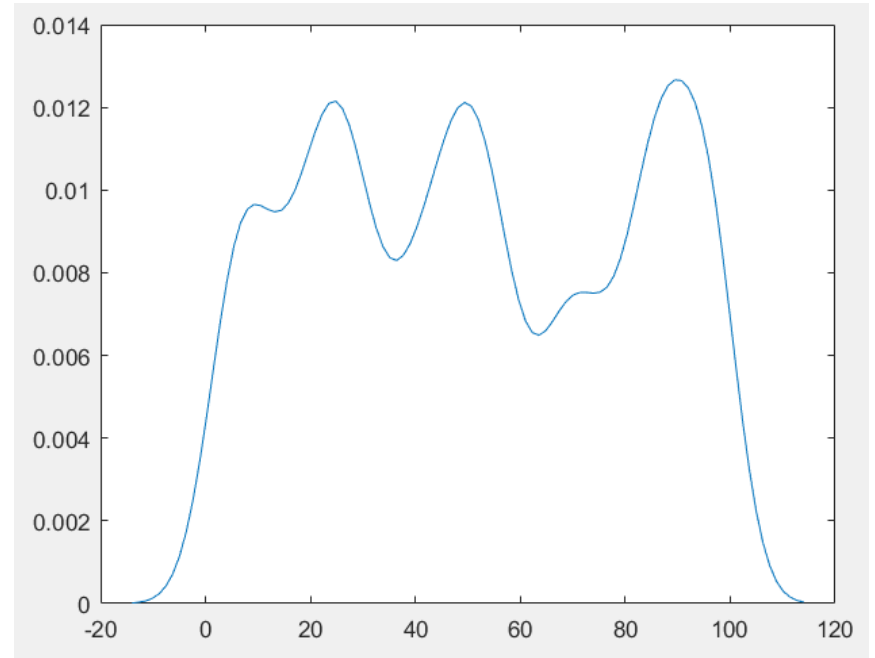
Kernel Estimator

- $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$
 - h : window width (smoothing parameter, bandwidth)
 - K is a kernel function
 - $\int_{-\infty}^{\infty} K(x)dx = 1$
- Intuitively,
 - Place a bump at each observation
 - Kernel estimator is a sum of bumps

Kernel Estimator



$h=1$



$h=5$

Kernel Estimator

- $\hat{f}(x)$ inherits all the continuity and differentiability properties of K
- e.g., when K is the normal density function, $\hat{f}(x)$ will be a smooth curve
- The parameter h controls the amount of smoothing

Nearest Neighbor Method

- Adapt the amount of smoothing to local density of data
- Generalized kth nearest neighbor density estimate
- $\hat{f}(t) = \frac{1}{nd_k(t)} \sum_{i=1}^n K\left(\frac{t-X_i}{d_k(t)}\right)$
- $d_k(t)$: kth nearest distance from the query point, t

Variable Kernel Method

- $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{hd_{i,k}} K\left(\frac{x-X_i}{hd_{i,k}}\right)$
- $d_{i,k}$: distance from X_i to the k th nearest point
 - Note that the window width is independent from query points
- The window width of the kernel placed on the point X_i is proportional to $d_{i,k}$
- It considers local density of observed data

General Weight Function

- $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n w(X_i, t)$
 - $\int_{-\infty}^{\infty} w(x, y) dy = 1$
 - $w(x, y) \geq 0$ for all x and y
 - e.g. histogram
 - $w(x, y) = \frac{1}{h(x)}$ if x and y are in the same bin
 - $w(x, y) = 0$ otherwise

Discussion

- Choice of Kernels

- Epanechnikov

- $K(t) = \frac{0.75\left(1 - \frac{1}{5t^2}\right)}{\sqrt{5}}$ for $|t| < \sqrt{5}$

- $K(t) = 0$ otherwise

- Biweight

- $K(t) = \frac{15}{16}(1 - t^2)^2$ for $|t| < 1$

- $K(t) = 0$ otherwise

- Gaussian

- $K(t) = \frac{1}{\sqrt{2\pi}}e^{-0.5t^2}$

Discussion

- Choice of Smoothing Parameters
 - Manual choices: visualize density estimation results with different parameters and select a proper one
 - Automatic choices: select optimal parameters that minimize errors

Error Estimation for Density Estimation

- Mean Squared Error (MSE) at a single point, t
 - $MSE_t(\hat{f}) = E \left(\hat{f}(t) - f(t) \right)^2$
 - $MSE_t(\hat{f}) = \left(E \left(\hat{f}(t) \right) - f(t) \right)^2 + V \left(\hat{f}(t) \right)$
 - $E \left(\hat{f}(t) \right) = \frac{1}{n} \sum E(w(X_i, t)) = \int w(x, t) f(x) dx$ (Silverman 1986, 36pp)
 - $V \left(\hat{f}(t) \right) = \frac{1}{n} V(w(X_i, t))$
 - $= \frac{1}{n} \left[\int w(x, t)^2 f(x) dx - \left\{ \int w(x, t) f(x) dx \right\}^2 \right]$
 - The bias, $E \left(\hat{f}(t) \right) - f(t)$, does not depend on the sample size, n
 - The variance, $V \left(\hat{f}(t) \right)$, decreases as we use more samples

Error Estimation for Density Estimation

- Asymptotic version (Silverman 1986, 39pp) for univariate data
 - $bias_h(x) = E(\hat{f}(x)) - f(x) \approx \frac{1}{2}h^2 f''(x)k_2$ (derived using Taylor expansion)
 - $V(\hat{f}(x)) \approx n^{-1}h^{-1}f(x) \int K(t)^2 dt$
- Trade-off between bias and variance
- In practice,
 - We need a way to estimate unknown terms $f(x)$, $f''(x)$

Discussion

- Radiance estimation in (progressive) photon mapping is an application of the density estimation
- The radius of the radiance estimation is related to the error terms
- Q. Can we choose an optimal radius in a data-driven way?