QFRM中心案例（二）：基于期权价格的金融资产未来价格分布的估计方法和实证

黄跃、刘依凡、钱天赐、李婷、黄勉*

上海财经大学数量金融与风险管理中心

摘　要

金融资产未来价格分布又称状态价格密度（State Price Density，简称SPD），是一种金融风险度量指标，能够反映投资者对未来资产价格分布的预期。虽然金融资产价格分布无法直接观测，但可以通过其期权衍生品的当前价格求解出来。近年来随着中国期权市场日益壮大，对金融资产价格分布估计的需求也在不断增加。本项目基于实时期权合约数据，同时使用看涨期权和看跌期权当前价格，综合运用核密度估计、局部回归、拒绝抽样法等非参数估计方法，得到期权相应标的资产未来价格分布的实时估计，并通过软件进行实时展现，为投资者衡量资产价格、进行市场交易提供一定的参考。虽然估计金融资产未来价格分布具有重要的实用价值，但当前几乎没有平台提供金融资产未来价格分布的实时展示。因此，本项目填补了市场空缺，是一个全新且具有实际意义的创新性工作。

关键字：金融资产未来价格分布、SPD、核密度估计、局部回归、拒绝抽样法。

*项目负责人，上海财经大学统计与管理学院。邮箱：huang.mian@mail.shufe.edu.cn；电话：021-65901057。

1 案例目标

金融资产未来价格分布又称状态价格密度（State Price Density，简称SPD），是一种反映投资者对金融资产未来价格分布预期的金融风险度量指标，能够为投资者和投资机构提供资产价格预期的参考。金融资产未来价格分布不能直接观测到，但可以通过该资产衍生品的当前价格导出。未来分布的估计依据当前时刻的市场情绪，衡量了资产价格未来出现在不同位置的可能性。相比于目前市场上常用的风险度量指标，金融资产价格未来分布不仅提供了市场的波动信息，还给出了具体的价格分布预期，可广泛应用于风险管理、波动预测等金融领域。

以欧式看涨期权合约为例，假设该合约标的资产价格为S，到期日为T，执行价为K。则该期权赋予持有者在T时刻以价格K买入该标的资产的权利。若T时刻标的资产价格$S_T > K$，则持有者行权，若$S_T < K$，则持有者不行权。因此持有看涨期权的到期收益函数为：

$$max(0, S_T - K) = (S_T - K)_+.$$

我们可以通过收益的数学期望及折现来给期权定价，其原理是公允价值应当等于到期时的收益期望。假设我们知道到期时资产价格S_T的分布$f(s)$，则当前时刻t，执行价为K的看涨期权价格应为

$$C_t(K) = e^{-r(T-t)} \int_K^{\infty} (s-K) f(s) ds, \tag{1}$$

其中r是无风险利率。

在公式(1)中，未来分布$f(s)$是不能观测到的，但t时刻不同执行价下的期权当前价格$C_t(K)$能够观测。因此我们可以通过期权价格反推$f(s)$，得到的就是S_T分布的估计，即标的资产的未来价格分布，反映了当前交易者对到期时标的资产价格的预期。

基于公式(1)，经过简单的微积分推导，可以得到金融资产未来价格分布的估计公式为

$$f(S_T) = e^{r(T-t)} \frac{\partial^2 C(K)}{\partial K^2}\bigg|_{K=S_T}, \tag{2}$$

即我们可以把期权价格作为执行价的函数，对该函数求二阶导，得到标的资产的未来价格分布估计。本项目基于ETF期权合约市场实时数据，综合使用局部回归法、拒绝抽样法、核密度估计等非参数估计方法，结合看涨期权和看跌期权信息，得到ETF的未来价格分布估计。未来价格分布估计为投资者提供合理的金融资产价格未来分布预期，并可实时展示，为投资者提供更丰富全面的风险参考依据。

2 案例背景

目前，中国金融衍生品市场受到重视程度增加，期权市场日渐发展壮大。2015年推出国内首支场内期权上证50ETF后，豆粕、白糖、铜期货期权相继上市，自2019年，又有多个指数期权和期货期权上市。随着期权市场的日渐完善，市场交易者的日渐增多，越来越多的投资者和投资机构希望通过实时更新的金融指标及时掌握市场动向、控制风险和收益，以便更好地参与市场交易，因此对相关金融风险指标的需求也随之增加，

对于未来资产价格，常用波动率指数（Volatility Index，VIX）来反应它的波动情况。但是，VIX仅仅只能反应未来价格分布的部分信息，而实际中往往不只关心未来价格的波动程度，还希望对价格能有更全面的了解，比如了解未来价格在各个位置取值的可能性。因此，对金融资产未来价格分布进行估计就很有必要。一方面，资产未来价格分布能够反映未来资产在各个价格上的密度，比VIX更全面地反应未来价格情况，能够为投资者衡量资产价格、进行市场交易提供更多的参考价值；另一方面，未来价格分布作为一种金融风险度量指标，也可以应用于金融分析和实践中，如风险管理、市场择时、交易策略设计等。

金融资产未来价格分布的估计方法包括参数方法和非参方法，参数方法基于资产价格分布和模型的假设条件，非参方法则不对模型做特殊要求。参数方法对期权标的资产价格过程进行模型假设，如基于Black & Scholes (1973)提出的BS模型，Heston (1993)提出的随机波动率模型等。但这些模型下的定价估计过于依赖模型假设，在实际中往往也很难验证假设的真实性，在假设不满足时无法保证参数方法的估计效果，具有较大局限性。非参数方法不施加模型假设，在一定程度上能很好地避免模型依赖问题，在估计中具有更大的灵活性。Aït-Sahalia & Lo (1998)首先提出一种基于局部多项式回归的非参数估计资产价格未来分布的方法，Aït-Sahalia & Duarte (2003)在之前模型的基础上增加了单调约束和凸约束，提高了模型的预测效果。

虽然中国期权市场日益壮大，对金融资产价格分布估计的需求也在不断增加，但目前国内几乎没有平台实时发布金融资产未来价格分布的估计来为投资者提供更多的信息。本项目使用非参方法估计金融资产未来价格分布：结合上证50ETF看涨期权和看跌期权的合约信息，首先使用局部回归给出金融资产未来价格的初步估计，再使用拒绝抽样法和核密度估计增加密度函数的合理性，得到修正的金融资产未来价格分布估计，并提供软件进行实时展示。

3 金融资产未来价格分布的估计方法

金融资产未来价格分布使用非参数方法进行估计。需要用到核密度估计、局部回归、拒绝抽样法，并根据期权合约相应数据计算得到。本项目基于上证50ETF给出标的资产未来价格分布的实时估计结果，具体估计方法细节如下。

3.1 估计方案细则

本项目基于上证50ETF期权合约给出标的资产的未来价格分布估计，称为上证50ETF未来价格分布。期权合约价格的读取规则和交易日读取时间段与QFRM-SVIX指数一致，详见QFRM中心案例（一）[5]。上证50ETF未来价格分布实时计算并展示最新的结果。在交易日规定的读取时间段内，每隔30秒进行一次更新；收盘及非交易日展示最近一次更新结果。

3.2 估计方法与步骤

使用非参数的方法估计金融资产未来价格分布。用到核密度估计、局部回归和拒绝抽样法，详细的方法说明见附录A。具体的估计步骤如下：

（1）分别对同一到期日的看涨期权合约和看跌期权合约使用局部回归估计，根据附录A中的公式(10)得到期权现价对执行价的二阶导，由公式(2)，局部回归结果可作为金融资产未来价格分布的初步估计；

（2）当标的资产预期到期价格高于现价时，使用基于看涨期权的局部回归估计部分；当标的资产预期到期价格低于现价时，使用基于看跌期权的局部回归估计部分。进行连续性修正后得到同时基于看涨期权和看跌期权的局部回归估计；

（3）由于局部回归结果不能保证函数非负且积分为1，使用拒绝抽样法从初步估计的分布中抽样，附录A中的公式(11)确保新的样本来自一个非负且积分为1的概率密度函数；

（4）根据附录A中的公式(3)对拒绝抽样法得到的样本进行核密度估计，得到完整的金融资产未来价格分布；

（5）分别对不同到期日的期权合约数据应用上述过程，得到标的资产价格不同时间的未来价格分布。

4 金融资产未来价格分布实证分析

本文实证基于上证50ETF期权数据，说明上证50ETF未来价格分布的估计过程。

4.1 数据来源及说明

以上证50ETF期权2021年3月26日的收盘数据为例。获取3月26日标的资产收盘价为3.514，并分别获取当日4个期权合约（4、5、6、9月）的行权价和看涨期权收盘价、看跌期权收盘价。

以4月到期的期权合约为例说明金融资产未来价格分布的计算过程。4月期权2021年3月26日距离到期日33天，行权价K和对应的看涨期权收盘价$C_t(K)$、看跌期权收盘价$P_t(K)$见表1。

表 1: 上证50ETF 4月期权2021年3月26日收盘价

K	$C_t(K)$	$P_t(K)$
3.1	0.41305	0.00375
3.2	0.31490	0.00725
3.3	0.22630	0.01635
3.4	0.14590	0.03640
3.5	0.08245	0.07375
3.6	0.04295	0.13320
3.7	0.01975	0.20900
3.8	0.00915	0.29830
3.9	0.00495	0.39435
4.0	0.00265	0.49500
4.1	0.00190	0.59210
4.2	0.00155	0.69490

执行价与期权价格的关系见图1，蓝色'o'代表看涨期权价格，橙色'x'代表看跌期权价格。

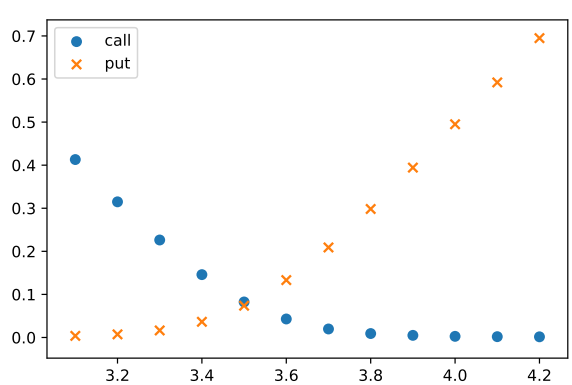
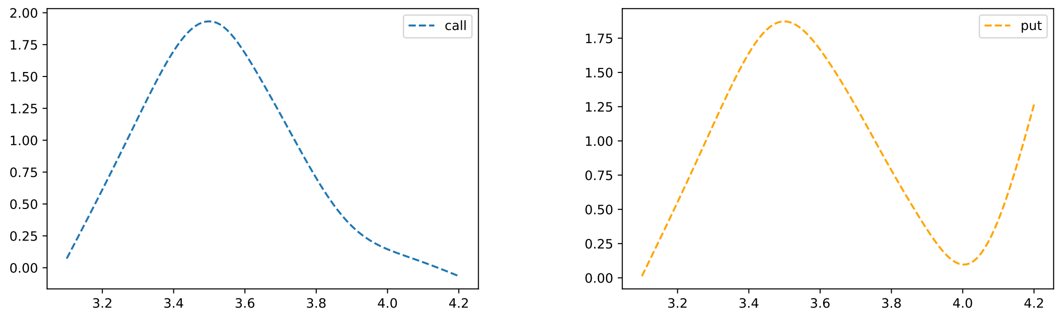


图 1: 期权价格散点图

4.2 基于4月期权数据的金融资产未来价格分布

根据行权价和对应的看涨期权价格、看跌期权价格获得期权价格的局部二阶估计，作为资产未来价格分布的初步近似，基于看涨期权和看跌期权的局部估计结果见图2。



(a) 看涨期权的局部回归估计

(b) 看跌期权的局部回归估计

图 2: 看涨期权和看跌期权的局部回归估计

当标的资产预期到期价格高于现价，即$S_T > S_t$ 时，使用基于看涨期权的估计部分；同理，当标的资产预期到期价格低于现价，即$S_T < S_t$时，使用基于看跌期权的估计部分。连续性修正后得到新的局部二阶估计，如图3所示，左边橙色部分基于看跌期权局部估计，右边蓝色部分基于看涨期权局部估计。

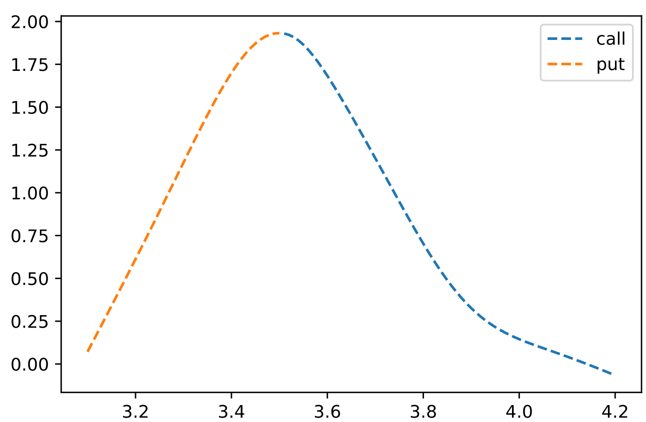


图 3: 修正后的局部回归估计

在修正后的局部回归结果的基础上使用拒绝抽样法进行抽样，进一步使用核密度估计，得到完整的基于4月期权数据的金融资产未来价格分布，展示在图4中。该估计分布展示了当前时刻市场投资者对于基底资产33天后价格分布的预期，横轴为基底资产价格，纵轴为基底资产33天后达到该价格的概率。其中红色虚线所在部分标出了估计的最大概率出现位置，在3.5处，比标的资产的现价3.514略低，说明投资者认为在33天后，标的资产价格最有可能出现在3.5处，分布略微右偏，市场总体有轻微的看跌倾向。

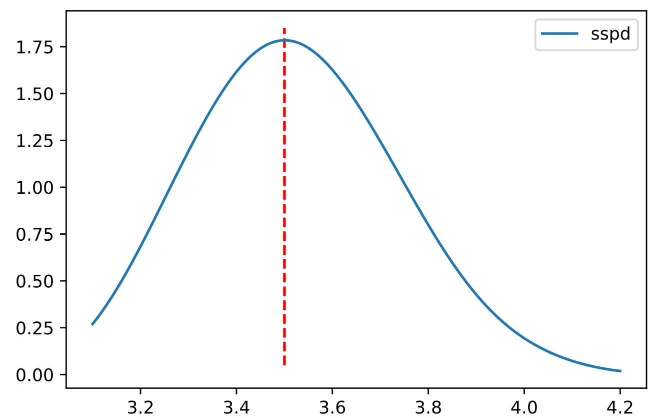


图 4: 基于4月期权数据的金融资产未来价格分布

4.3 基于不同到期日期权数据的金融资产未来价格分布

分别对4月、5月、6月、9月期权数据重复上述计算过程，得到基于不同到期日期权数据的金融资产未来价格分布，共同展示在图5中。

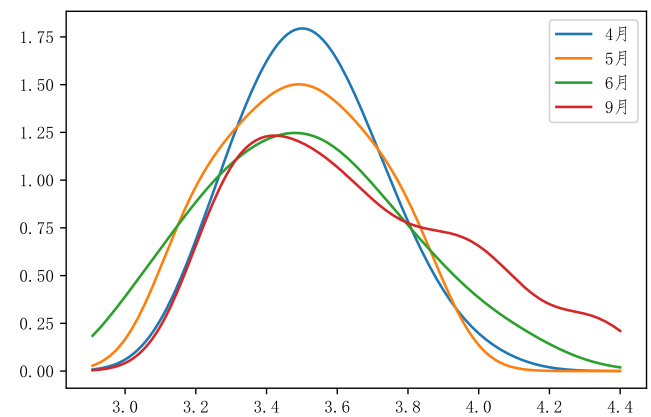


图 5: 基于不同到期日期权数据的金融资产未来价格分布

此时4月、5月、6月、9月期权合约剩余到期日分别为33天、61天、89天、180天。因此不同的金融资产未来价格分布分别反映了投资者对标的资产未来33天、61天、89天、180天的价格分布期望。从图5可以看到，随着到期时间的增加，金融资产未来价格分布最大概率出现的位置逐渐左移，说明目前投资者对市场持有轻微的看跌态度。还可以看到，随着到期时间的增加，未来价格分布的分散性逐渐增加，政策、突发事件、假期等未知影响因素使资产价格分布的不确定性增加，市场投资者对资产价格的看法也会更为不同。估计中出现局部极值，也可以参考价格的支撑位/阻力位进行进一步的分析。

5 总结与展望

金融资产未来价格分布能够通过金融资产衍生品的现价估计投资者对该金融资产价格的未来预期。本项目以50ETF为标的资产，以50ETF期权合约为基础估计上证50ETF的未来价格分布。综合应用核密度估计、局部回归二项估计、拒绝抽样法，结合看涨期权与看跌期权的估计结果，得到资产未来价格分布的非参数估计，直观地反映了当前时刻投资者对基底资产未来价格的看法。实证表明，金融资产未来价格分布具有合理性和参考性。

局部回归二项估计作为标的资产未来价格分布的初步估计具有理论优势，在局部回归的基础上综合了看涨期权和看跌期权的估计结果，增加了平滑性处理、拒绝抽样法和核密度估计的修正，使得估出的金融资产价格分布更为合理，充分利用了交易期权的信息。

通过不同到期日的期权数据得到的金融资产未来价格分布能够反映金融标的资产在未来不同时期的价格分布预期。相比于VIX等波动率金融风险指标，金融资产未来价格分布不仅能反映市场的波动程度，还能够反映资产价格在各个位置的概率预期，从而为投资者提供更为全面的参考建议。

本项目中的金融资产未来价格分布估计在非负性、光滑性、合理性上表现良好，客观地反映了投资者对未来资产价格预期的可能性，能够为投资者和投资机构提供投资建议和参考。本项目未来计划在设计细节上做更多考量，并在更多类型的衍生品上估计标的金融资产的未来价格分布。

6 软件产品描述

中心研究团队自主开发的软件产品VIXClient上展示了上证50ETF未来价格分布。交易日内，每隔30秒实时更新一次上证50ETF未来价格分布；非交易日，展示最近一次交易日的更新结果。同时展示四个不同到期日的期权合约对应的上证50ETF未来价格分布，表示投资者对不同时间的资产价格预期。投资者可以使用VIXClient 软件随时查看上证50ETF未来价格分布，作为市场风险的参考指标。

VIXClient下载链接：ds.shufe.edu.cn/qfrm/producttool.html

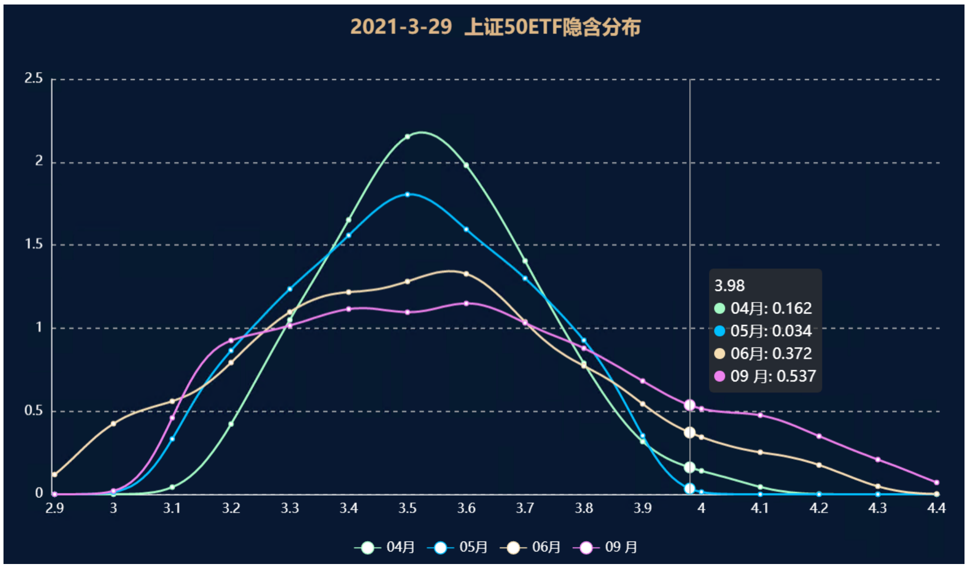


图 6: 上证50ETF未来价格分布界面展示

A 附录

A.1 核密度估计

核密度估计是一种用于估计数据概率密度的非参数平滑方法，并且对于维数较低的数据具有比较好的可视化效果。

假设$x_1, x_2, \cdots, x_N$是独立同分布的一维数据样本，服从概率密度函数$f(x)$，则$f(x)$ 的核密度估计为

$$\hat{f}_h(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(x_i - x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x_i - x}{h}\right), \tag{3}$$

其中，$K_h(\cdot) = h^{-1}K(\cdot/h)$中的$K(\cdot)$是一个均值为0的密度函数，称为核函数。$h > 0$ 是一个平滑参数，也称为窗宽。由于$K_h(\cdot)$是密度函数，满足非负、连续、积分为一的条件，由此得到的核密度估计$\hat{f}_h(x)$与将满足非负、连续、积分为一，因此是一个密度函数。

在核密度估计中，需要先确定核函数的类型及窗宽的大小。

核函数有很多类型可供选择，常见的有均匀分布核函数、高斯分布核函数、Epanechnikov核函数等。在金融资产未来价格分布中，我们使用较为常用的高斯分布核函数，假设核函数的变量为u，则公式为

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \tag{4}$$

相比核函数类型，窗宽h的选择更为重要。窗宽反映了核密度估计对各个数据点使用的比重。窗宽小说明核密度估计用了x附近少量的数据，因此估计曲线相对不够平滑，属于偏差较小方差较大的估计；窗宽大说明核估计用了x附近更多的数据，因此估计曲线更平滑，属于偏差较大的估计。在高斯核函数中，h恰好表示正态密度核函数的标准差，通常用以下准则选取核密度估计的最优窗宽:

$$h_{opt} = C * \hat{\sigma} N^{-\frac{1}{5}}, \tag{5}$$

其中，C为调节系数，在高斯核中设置为1.06，$\hat{\sigma}$为样本标准差，N为样本个数。

A.2 局部回归

局部回归是非参估计中重要的估计方法。局部多项式回归首先对x邻域内的函数f使用泰勒展开,

$$f(X) \approx \beta_0 + \beta_1(X-x) + \cdots + \beta_p(X-x)^p. \tag{6}$$

最小化如下目标函数获得$\beta_0, \beta_1, \cdots, \beta_p$的估计，相当于使用核函数作为权重的加权最小二乘法,

$$\sum_{i=1}^{N} \left[y_i - \beta_0 - \beta_1(X-x) - \cdots - \beta_p(X-x)^p\right]^2 K_h(x_i - x), \tag{7}$$

则f局部估计的显示表达解为

$$\hat{\beta}(x) = (X^T W X)^{-1} X^T W y, \tag{8}$$

$$\hat{f}(x_0) = e_1^T (X^T W X)^{-1} X^T W y, \tag{9}$$

其中,

$$W = diag[K_h(x_1 - x), \cdots, K_h(x_N - x)],$$

$$e_1 = (1, 0, 0, \cdots, 0)^T,$$

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_2^T \end{pmatrix} = \begin{pmatrix} 1 & x - x_1 & \cdots & (x-x_1)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x - x_n & \cdots & (x-x_n)^p \end{pmatrix}.$$

根据计算结果，也可以得到f在x处的各阶导数估计。$f(x)$在x处的$q(q < p)$阶导数估计为

$$\hat{f}^{(q)}(x_0) = q! e_{q+1}^T (X^T W X)^{-1} X^T W y, \tag{10}$$

其中e_{q+1}为第$q+1$个元素为1，其他元素均为0的单位向量。在金融资产未来价格分布估计中，我们利用局部回归对价格密度做初步估计，在公式（10）中取$q = 2$。

A.3 拒绝抽样法

拒绝抽样又称为接受-拒绝抽样，是一种简单高效的抽样方法，可以通过抽样从已知分布中产生一个概率密度函数。

考虑从某个分布$f(x)$中抽样，$f(x)$形式已知，但不一定满足积分为1，也就是说，假设$f(x) = ch(x)$，$h(x)$已知但c可以是未知的。在实际问题中，$f(x)$是难以直接抽样的，我们从一个相对容易抽样的分布$q(x)$中产生样本，例如$q(x)$是正态分布或均匀分布。

如果存在一个整数M满足条件

$$\sup_x \frac{h(x)}{q(x)} < M < \infty,$$

则拒绝抽样法算法描述如下：

- 抽样$X \sim q$，$U \sim Uniform[0,1]$，
- 如果$U \leq \frac{h(X)}{Mq(X)}$，则接受$X$，否则拒绝$X$。

可以证明，从拒绝抽样算法中接受的样本服从分布$f(x)$。由于

$$P\left(X \leq x | X\ is\ accepted\right) = P\left(X \leq x | U \leq \frac{h(X)}{Mq(X)}\right) = \frac{P\left(X \leq x, U \leq \frac{h(X)}{Mq(X)}\right)}{P\left(U \leq \frac{h(X)}{Mq(X)}\right)},$$

其中，分母和分子可以分别化为

$$P\left(U \leq \frac{h(X)}{Mq(X)}\right) = E_q\left[P(U \leq \frac{h(X)}{Mq(X)})|X\right] = E_q\left[\frac{h(X)}{Mq(X)}|X\right] = \frac{1}{cM},$$

$$P\left(X \leq x, U \leq \frac{h(X)}{Mq(X)}\right) = \frac{\int_{-\infty}^{x} f(x)dx}{cM},$$

于是有

$$P\left(X \leq x | X\ is\ accepted\right) = P\left(X \leq x | U \leq \frac{h(X)}{Mq(X)}\right) = \int_{-\infty}^{x} f(x)dx. \tag{11}$$

拒绝抽样法能够保证样本分布来自非负且积分为1的密度函数。

参考文献

[1]. Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637-654.

[2]. Heston, S.L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327-343.

[3]. Aït-Sahalia, Y. and Lo., A.W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, 53, 499-547.

[4]. Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1), 9-47.

[5]. 上海财经大学数量金融与风险管理中心. (2021). QFRM中心案例（一）：SVIX市场波动指数的设计、实证和软件产品.